1996

Modeling the elevation characteristics of the head-related impulse response

C. Phillip Brown
San Jose State University

MODELING THE ELEVATION CHARACTERISTICS OF THE HEAD-RELATED IMPULSE RESPONSE

A Thesis

Presented to

The Faculty of the Department of Electrical Engineering

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

C. Phillip Brown

May 1996

APPROVED FOR THE DEPARTMENT OF

ELECTRICAL ENGINEERING

Dr. Richard Duda

Dr. Benjamin Knapp

Dr. Kevin Jordan (Department of Psychology)

APPROVED FOR THE UNIVERSITY

ABSTRACT

MODELING THE ELEVATION CHARACTERISTICS OF THE HEAD-RELATED IMPULSE RESPONSE

by C. Phillip Brown

This thesis presents the research performed to develop and validate a simple signal-processing-based model of the head-related impulse response (HRIR). The model captures elevation as well as azimuth cues. The simplicity of the model permits efficient implementation in signal processing hardware, allowing for real-time operation. The parameters in the model can be adjusted to fit a particular individual's HRIR. The evaluation is based on listening tests in which the output of the model is compared to that of experimentally measured HRIR's.

Acknowledgments

I would like to thank Dr. Richard O. Duda of the San Jose State University Electrical Engineering Department for his guidance, patience and wisdom over the past year. Without his support, none of this would have been possible.

Both Nathan Henderson and Dr. Duda are owed a great deal of thanks for having the patience to listen to long sessions of white noise bursts while I worriedly hovered over them. I also thank Nathan for stopping in the middle of his thesis work whenever I said "Hey, come over here and listen to this!"

I would like to thank Richard Lyon of the Advanced Technology Group at Apple Computer for his insight into the head-shadow model and his support of the Digital Signal Processing lab at SJSU.

I would also like to thank the members of my thesis committee for taking time out of their busy schedules to review my work.

Finally, I would like to thank my wife Rebecca for her continued encouragement, support and affection.

This work was financially supported by the Interactive Systems Program of the National Science Foundation under grant number IRI-9402246.

Table of Contents

List of Tables

List of Figures

Nomenclature

ADC	analog to digital converter
DSP	digital signal processing
FIR	finite impulse response
HRIR	head-related impulse response
HRTF	head-related transfer function
Hz	Hertz (cycles per second)
IID	interaural intensity difference
IIR	infinite impulse response
ITD	interaural time difference
SNR	signal to noise ratio
θ	azimuth
ϕ	elevation
r	range
ρ_{pn}	pinna echo magnitude
τ_p	pinna echo delay
τ_{sh}	shoulder echo delay
μs	micro second
ms	milli second

Introduction and Motivation

The topic of spatial hearing in humans has been one of increasing interest over the past several years, while research into the subject has existed for literally hundreds of years. The recent introduction of virtual environments via computer modeling has fueled interest into creating three-dimensional (3D) sound sources, not only within the academic field, but in the private sector as well.

In order for a person to localize sound in three dimensions, the sound must arrive (or appear to have arrived) from a specific azimuth (θ), elevation (ϕ), and range (r). It is well established that the primary cues for localizing sounds in the horizontal plane (azimuth) are binaural: the interaural intensity difference (IID) and interaural time difference (ITD) [2]. Cues for localizing sound in the vertical plane (elevation) appear to be primarily monaural, although studies have shown that elevation information can be recovered from IID alone [5]. The cues for range are the least understood, and are typically associated with room reverberation.

This thesis concentrates on identifying the elevation-dependent components of the head-related impulse response (HRIR). The HRIR, which varies with both azimuth and elevation angle, captures the physical effects of the diffraction of sound waves by the torso, shoulders, head and pinnae.

It is known that elevation cues can be created synthetically by convolving a

measured HRIR response with a sound source, and then reproducing the convolved signal over headphones. This process lies at the heart of such devices as the Convolvotron [6]. Unfortunately, this approach relies heavily on tables of experimentally measured HRIR's. This is not only computationally complex, but requires individual HRIR measurements to get accurate elevation cues. A simple model, based on the elevation-dependent features found in measured HRIR data, would lead to efficient implementation in real-time signal processing.

Background

Significant research into the elevation characteristics of the HRIR has been made over the past thirty years. Batteau[1] made time-domain measurements on a model ear and proposed that the outer ear, or pinna (Figure 1), acts as a reflector which introduces delayed replications (i.e., echoes) of the incident sound. Blauert [2] and Wright et al. [14] have observed that similarities exist between the frequency response measurements they have made of the outer ear and the comb-filter effects of Batteau's echoes, although Blauret considers the concept of an "echo" as ill-defined when the structures are the same size as a wavelength. Watkins [12] confirmed that a model of two such echoes can produce elevation effects.

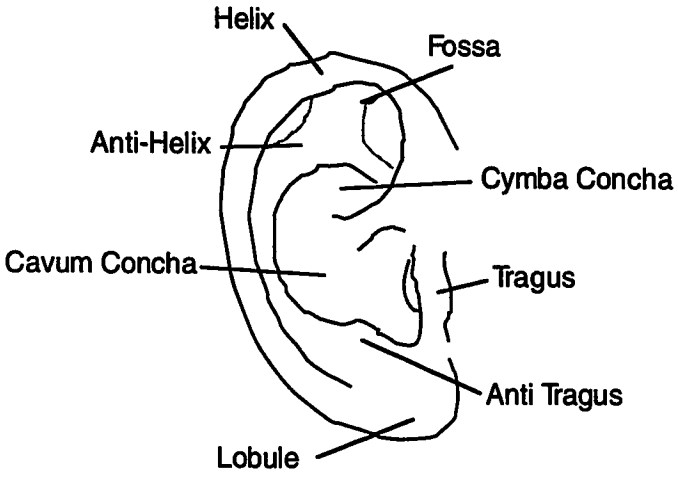


Figure 1 - Outer Ear (Pinna) and sub-components

Duda [4] has investigated the head-related transfer function (HRTF), which is the frequency-domain representation (magnitude and phase) of the HRIR. Spherical head models for the interaural intensity difference (IID) and interaural time difference (ITD) are discussed, as well as pinna echo models. Duda concludes that azimuth effects are readily modeled, while elevation effects remain difficult to model. The difficulty lies in mathematically determining the effects of human body geometry on wave propagation [11]. Structural modeling [7] may be a more effective method, provided the model parameters can be determined.

Richard Lyon of Apple Computer (private communication) has developed a "head-shadow" model that is similar to Duda's IID model. Figure A in Appendix A illustrates the inverse head-shadow frequency response at a variety of azimuth angles, for both the near (ipsilateral) and far (contralateral)

ears. The head-shadow model is a pole-zero filter that adjusts the IID based on frequency and azimuth. When combined with the correct ITD information, an effective azimuth localization is achieved. See Appendix A for more information.

A structural HRIR model that breaks down the various physical parameters of the body into components allows a more intuitive "block diagram" approach. Genuit [7,8] has proposed a model consisting of cascade and multi-path filters (Figure 2). The basis for the filters originate from Kirchhoff's solution to the wave equation for approximate body geometries. Genuit incorporates static features of the HRTF (ear-canal resonance and eardrum impedance), as well as azimuth-dependent (ITD, IID) and elevation-dependent (pinna and shoulder echo) features in the model. Unfortunately, the model suffers from limitations in that it approximates complicated human geometries with crude physical models. In addition, to the best of our knowledge, Genuit's model has never been validated.

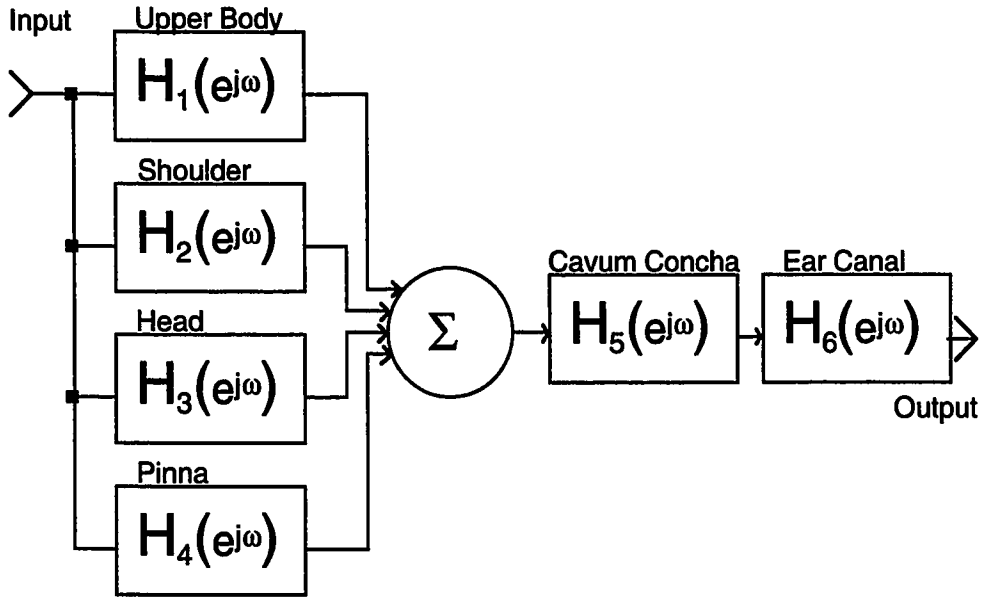


Figure 2 - Block Diagram of Genuit's Model

Cassaro and Van Belleghem [3] have implemented a digital-signal-processing (DSP) filter to synthesize binaural sounds. The filter contains head-shadow, ITD, shoulder and pinna effects. However, they did not have adequate data to evaluate the ability of such a model to recreate spatial effects.

Approach

The approach to developing the model described in this thesis is based in the time domain. Time-domain representation offers the advantage of preserving magnitude and phase information in a single, coherent signal. Because the phase response is difficult to work with, frequency-domain representation of the head-related transfer function (HRTF) often invites the researcher to concentrate the analysis on magnitude components and to put

less emphasis on the detailed phase (e.g., timing) information.

To model the elevation characteristics of the HRIR, three stages of research were performed. The first stage involved measuring and collecting raw HRIR data on three subjects. The second stage was the removal of the azimuth dependent features contained within the measured HRIR's. The third stage was modeling the elevation-dependent features that remain in the modified HRIR's.

The HRIR data was obtained using a modified Crystal River Engineering (CRE) Snapshot™ system on a Pentium-based PC. Subsequent signal processing and modeling were performed using Matlab™ on a Apple Power Macintosh 8100. Removal of the azimuth-dependent information (IID and ITD) from measured HRIR data allows the elevation-dependent features of the HRIR to surface. Once these features are identified, they are modeled using simple filtering techniques. The azimuth dependent effects are then re-introduced.

The model structure incorporates echoes generated by reflections of the outer ear (pinna) and shoulders. Since localization cues in the horizontal plane (azimuth) tend to be similar from person to person [13], the horizontal localization is based on modified versions of existing IID and ITD models [4].

The implementation of the model is based on simple filtering techniques. The head-shadow model is a recursive or infinite impulse response (IIR)

design. The shoulder and pinna echoes are modeled using a non-recursive or finite impulse response (FIR) approach. These simple filters are used in cascade and multi-path configurations to achieve a model that synthesizes azimuth and elevation cues (Figure 3).

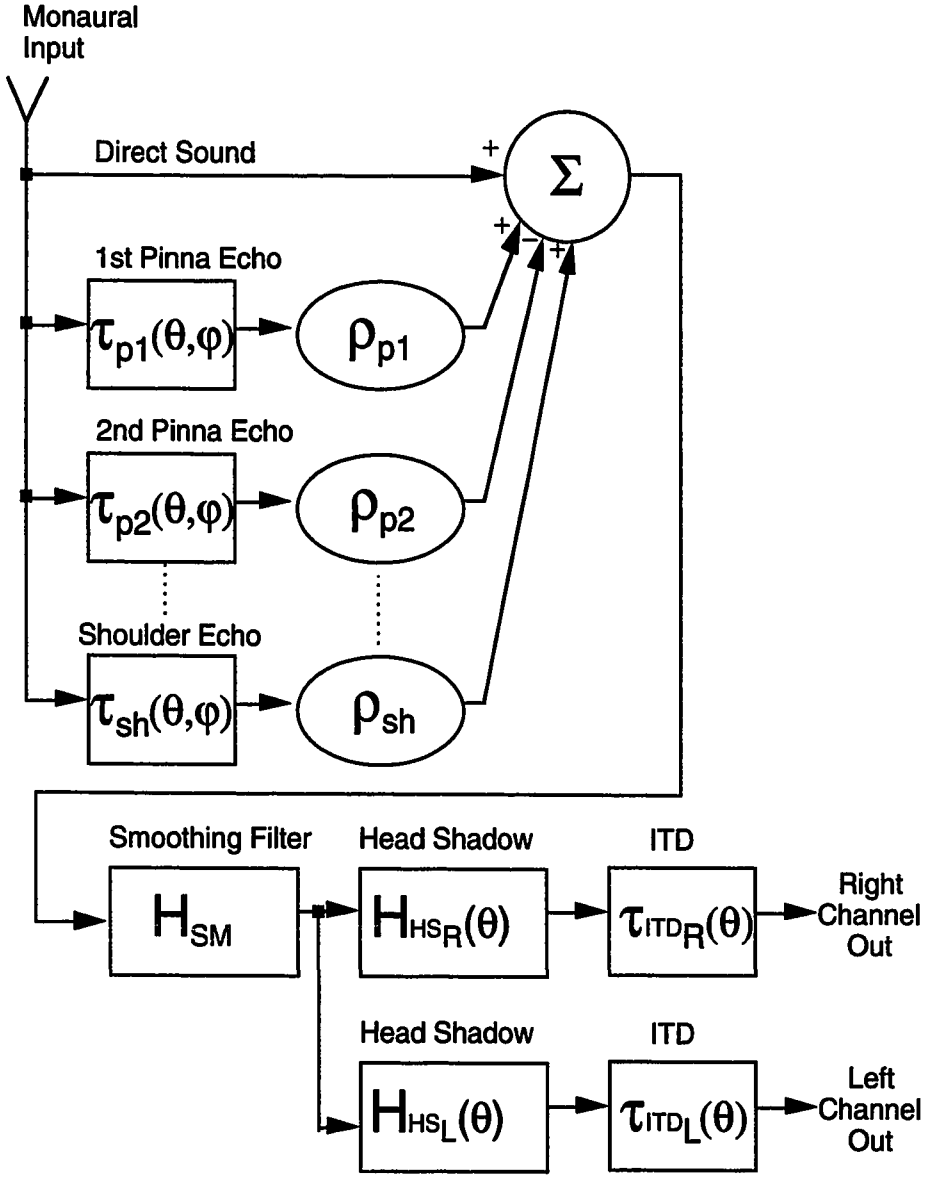


Figure 3 - HRIR Model Block Diagram

The optimization of the filter parameters was accomplished using standard system identification techniques and interactive modification. To validate the model, listening tests were performed through headphones and in-ear phones (which eliminate some of the artifacts that arise through headphone listening).

Measurements

HRIR measurements were made in the Engineering Building at San Jose State University (SJSU). The measurements were taken in an anechoic chamber (Industrial Acoustics Company, schedule 40, model 40) to minimize room acoustics which could affect the measurements. HRIR data was originally obtained using an stock (unmodified) Snapshot system. The Snapshot system uses Golay codes (maximum-length sequences) rather than actual impulses to measure the HRIR. The advantage of this approach over traditional measurement techniques is an increased signal-to-noise ratio (SNR). The Snapshot system also performs a minimum-phase reconstruction of the measured HRIR. During the minimum-phase reconstruction, the data is time-aligned and equalized for the low-frequency roll-off that occurs in the sound source (speaker). Unfortunately, while it does simplify some computations, the minimum-phase reconstruction changes the HRIR wave form, and might disrupt the timing cues. The data is finally stored as a stereo, 44.1 kHz, 16-bit, 128 sample representation of the HRIR.

The Snapshot system uses a 72-point cylindrical coordinate system, with 30° azimuth resolution and 18° elevation resolution. The azimuth range is a full $0° \le \theta \le 360°$; the elevation range is $-36° \le \phi \le +54°$. After examining the data, we decided that a finer degree of resolution was required to allow us to identify the elevation-dependent trends. A much finer elevation resolution of 5° was decided upon, and it was also decided to adopt a constant-ITD polar coordinate system (see Appendix B).

A fixture was designed to accommodate the polar coordinate system and the increased angular resolution (Appendix C). The fixture allowed vertical movement (elevation) from $-85° \le \phi \le +90°$ by adjusting a pivoting pole, with the pivot point along the subject's interaural axis. The range from the source to the subject's center-of-head was set to approximately 4 feet. The azimuth angle was set by moving the fixture with respect to the test subject, while keeping a constant range. The Snapshot software was then modified to accommodate the increased angular resolution and the polar coordinate system. Additionally, the software was changed to eliminate the minimum-phase reconstruction, at the expense of having no compensation for the low-frequency roll-off of the speaker.

Measurements were made on R. Duda, N. Henderson, and P. Brown at a constant azimuth angle of 55°, while the elevation angle varied in 5° increments from $-85° \le \phi \le +90°$. The subjects were seated on a small adjustable stool during the measurements, so that the subject's interaural axis was aligned with the pivot point on the fixture.

To validate the modified Snapshot system, data was also taken on subject PB using a Hewlett-Packard pulse generator to drive the sound source and a Macintosh II computer with a Digidesign 16-bit, 44.1 kHz, two-channel analog-to-digital converter (ADC) to record the signal. Nearly identical results were obtained with this system, when compared to the modified Snapshot system. All subsequent measurements were then obtained using the modified Snapshot system.

The raw Snapshot data was convolved with Gaussian white noise of duration 500 mS and played back over headphones. Through informal listening tests, it was determined that elevation cues exist in the measured HRIR data, although these cues differed between subjects.

Further measurements were then taken on P. Brown at constant azimuth angles of θ=0°, 15°, 30°, 45° and 60° and elevation angles ranging from -80°$\leq\phi\leq$+80° in 5° increments. To reduce the effects of head motion on these measurements, a laser-pen pointer was affixed to the subject's head and targeted on a fixed point on the wall. Listening tests also indicated that this data included elevation (as well as azimuth) cues. Because the subject was seated, significant torso shadowing was present in the median plane near -80°$\leq\phi\leq$-60°.

Signal Processing

The first step in processing the raw HRIR data was to remove the free-field response contributed by the sound source (Bose loudspeaker) and blocked-meatus (in-ear) microphones. While the free-field response has a "fixed" transfer function (e.g., not dependent on azimuth or elevation), its removal restores the data to a form that is independent of the measurement equipment. This was particularly important in restoring the low-frequency response which was rolled off due to the small size of the loudspeaker.

Outside of the median plane (θ=0°), the far-ear response quickly becomes shadowed by the diffraction of sound around the head. To visually compare the features of both the near and far ears (outside of the median plane), an inverse head-shadow filter was applied to the measured HRIR responses (Appendix A). At each azimuth angle, the HRIR's were filtered to remove the head-shadow effect at that particular azimuth. In the median plane the inverse head-shadow filter has no effect. As the azimuth angle increases towards 60°, the head-shadow effect increases, and therefore the compensation from the inverse head-shadow filter also increases. Removing the head-shadow effect restores the details that would otherwise be lost due to the large amplitude difference between the near and far ears.

Once the head-shadow effects were removed, the HRIR's were interpolated (using the Matlab function "interp.m") by a factor of four and then time aligned so that the impulse response begins at the first sample. The time

alignment removes the absolute timing difference between measurements due to head movement, minor mechanical inaccuracies, etc. The interpolation by four is critical in time-aligning the impulses as it greatly reduces timing "jitter" that occurs at the nominal sample rate. Once the responses are time-aligned, the elevation-dependent trends are more easily seen, and it is simple to compare the near- and far-ear differences. Next, it was necessary to determine the best way to examine this data, which was now time-aligned and had the head-shadow effects removed.

It was decided that a gray-scale plot provided the best visual representation of elevation-varying features of the HRIR. The gray-scale plot is read as follows: HRIR amplitude is represented by a gray level, as indicated by the color bar on the right of each plot. Time increases from top to bottom along the y-axis, and the elevation angle decreases from left to right along the x-axis.

The first set of gray-scale plots compares the far-ear to the near-ear responses at each azimuth (θ=0°, 15°, 30°, 45°, 60°), over the elevation range $-80° \leq \phi \leq +80°$ (Figure 4). The second set of plots (Figure 5) compares all of the near (or far) ear responses at the five azimuth angles. Figure 6 illustrates the schematized features we observed in these plots, named as "echoes" for ease of interpretation. The original signal, pinna echoes and shoulder/torso scattering are shown as discrete lines, with positive amplitude shown in white and negative amplitude in black.

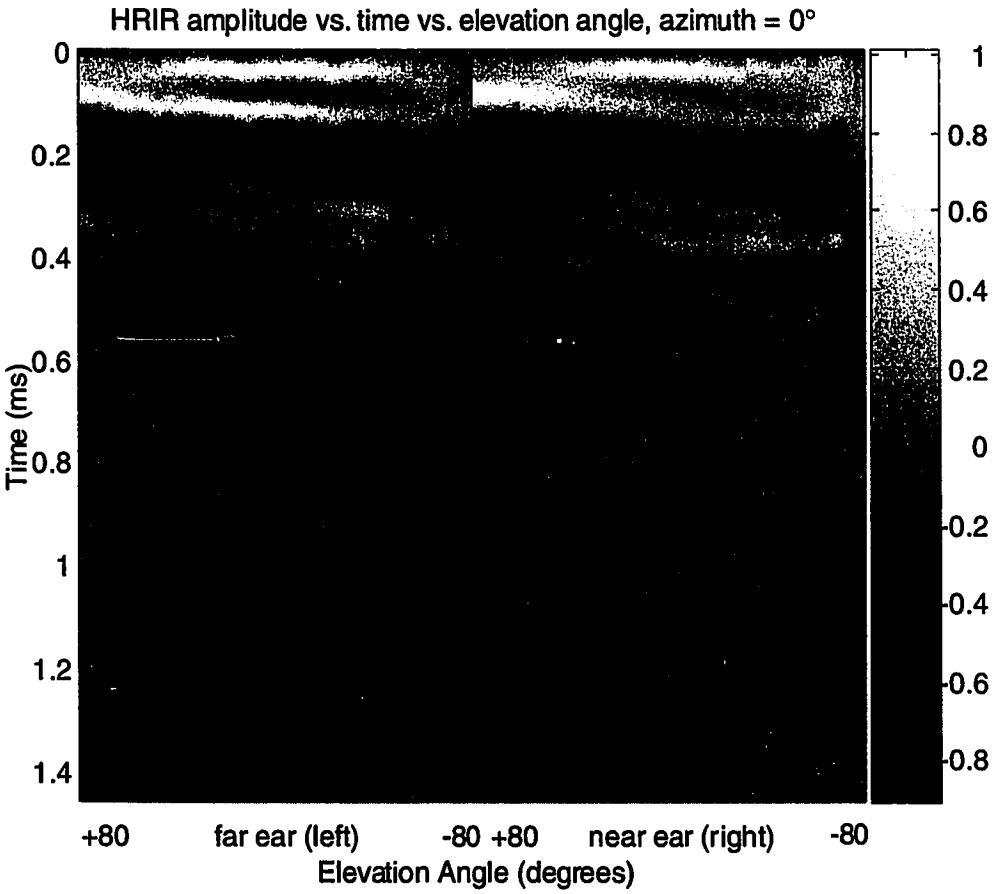


Figure 4a - HRIR Response for θ=0° (Subject PB)

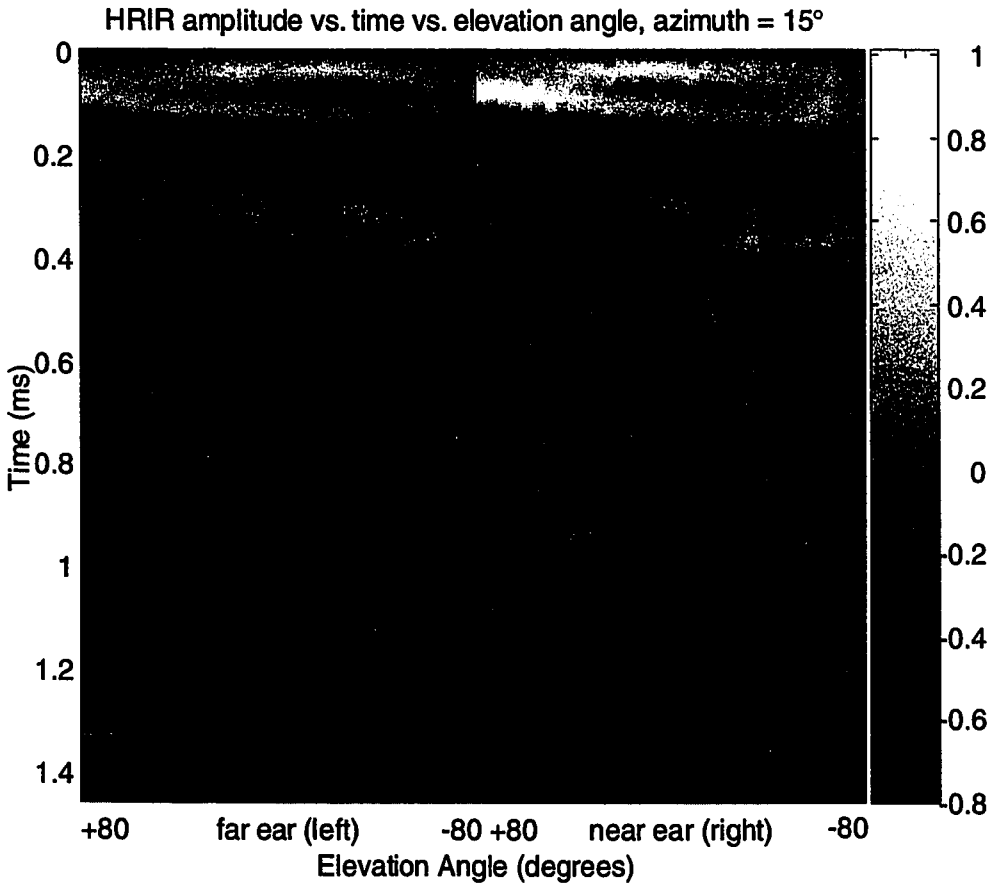


Figure 4b - HRIR Response for θ=15° (Subject PB)

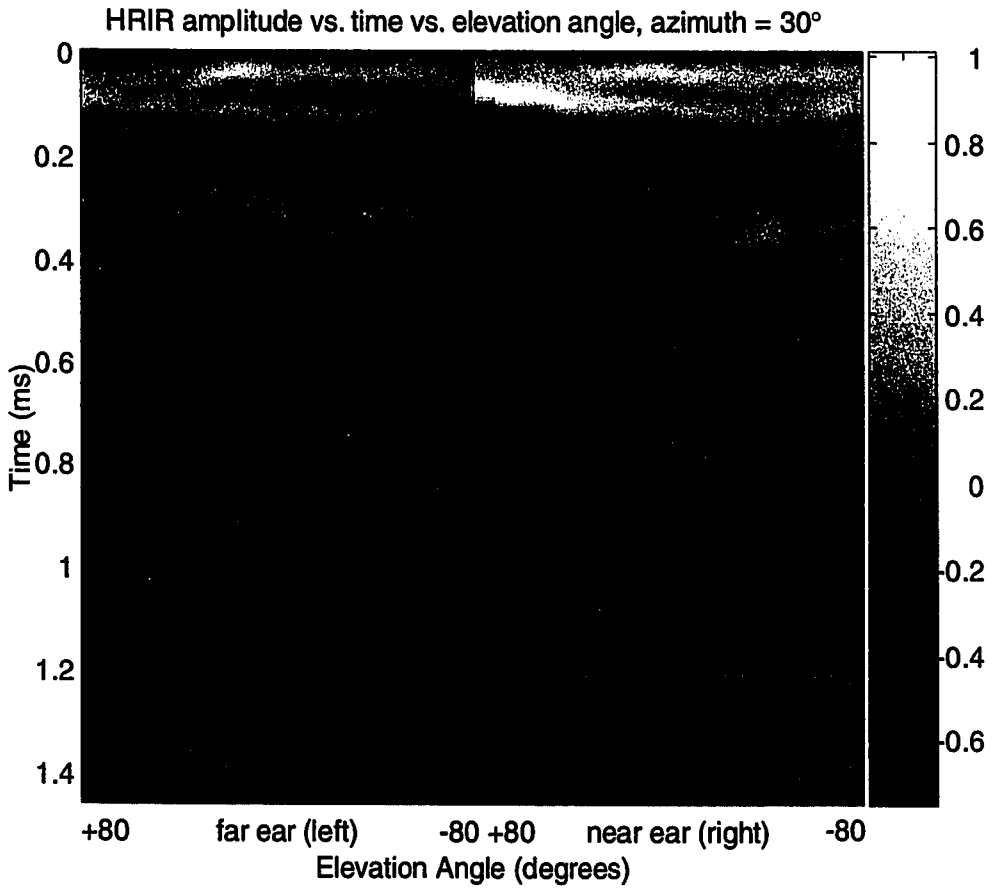


Figure 4c - HRIR Response for θ=30° (Subject PB)

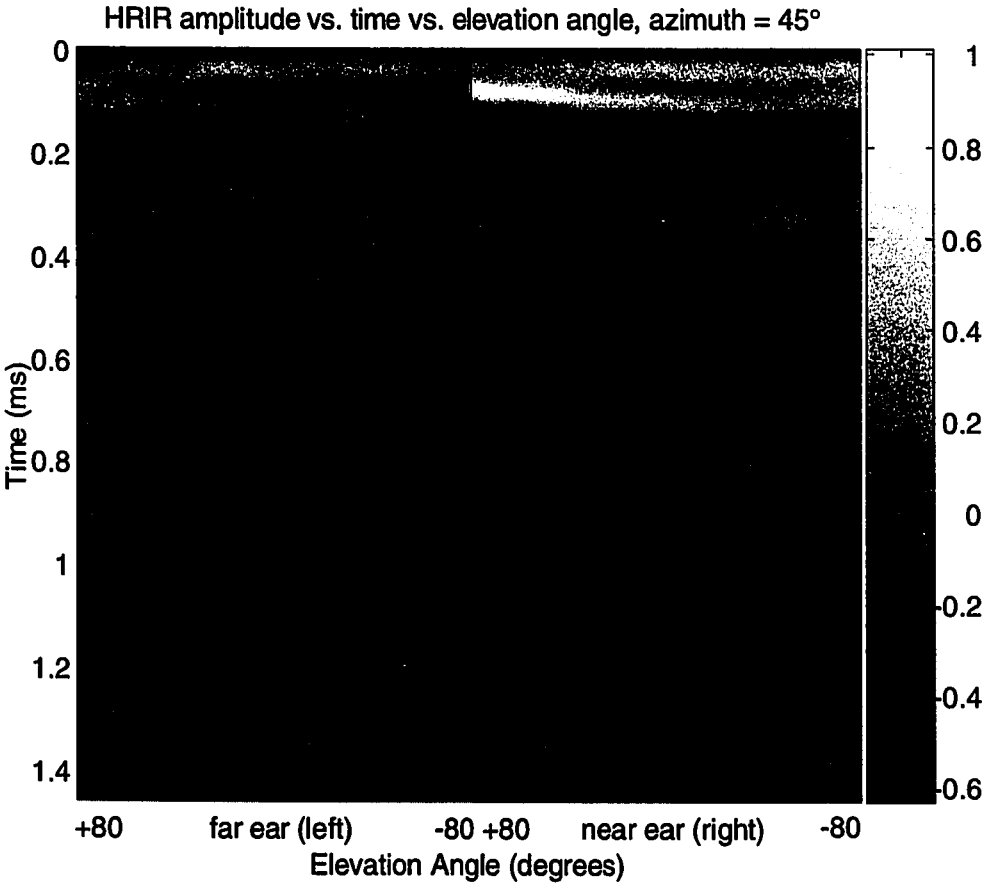


Figure 4d - HRIR Response for θ=45° (Subject PB)

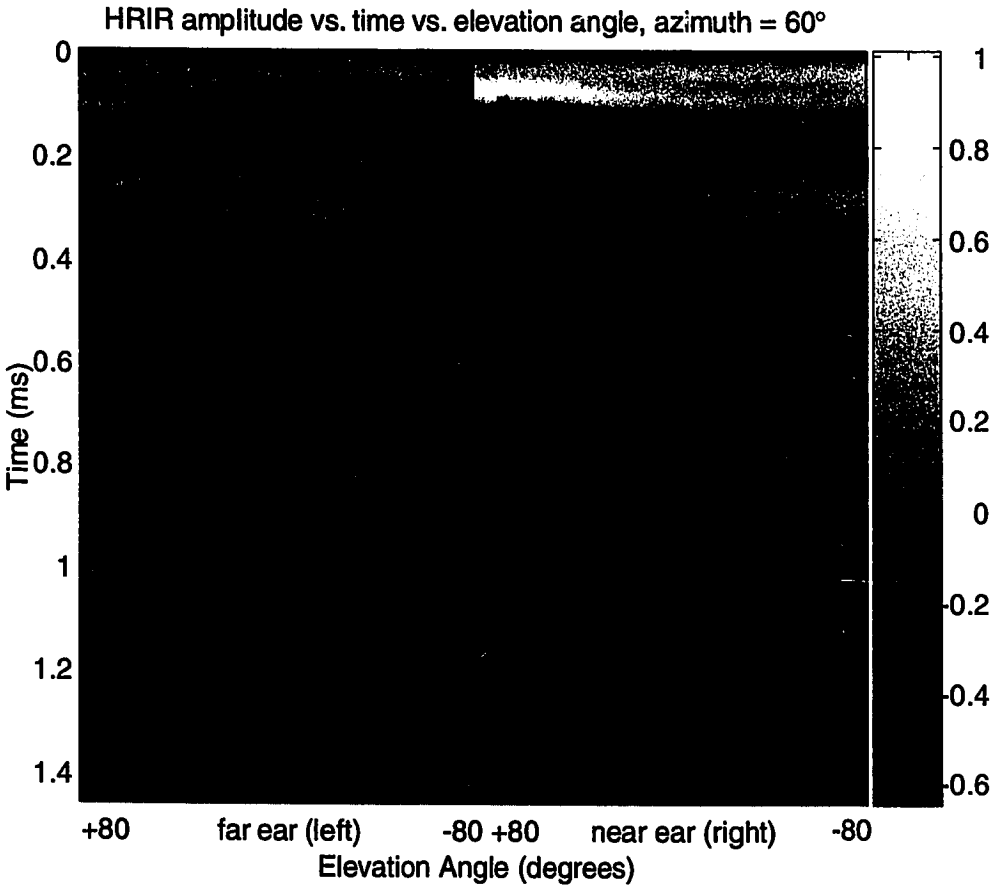


Figure 4e - HRIR Response for θ=60° (Subject PB)

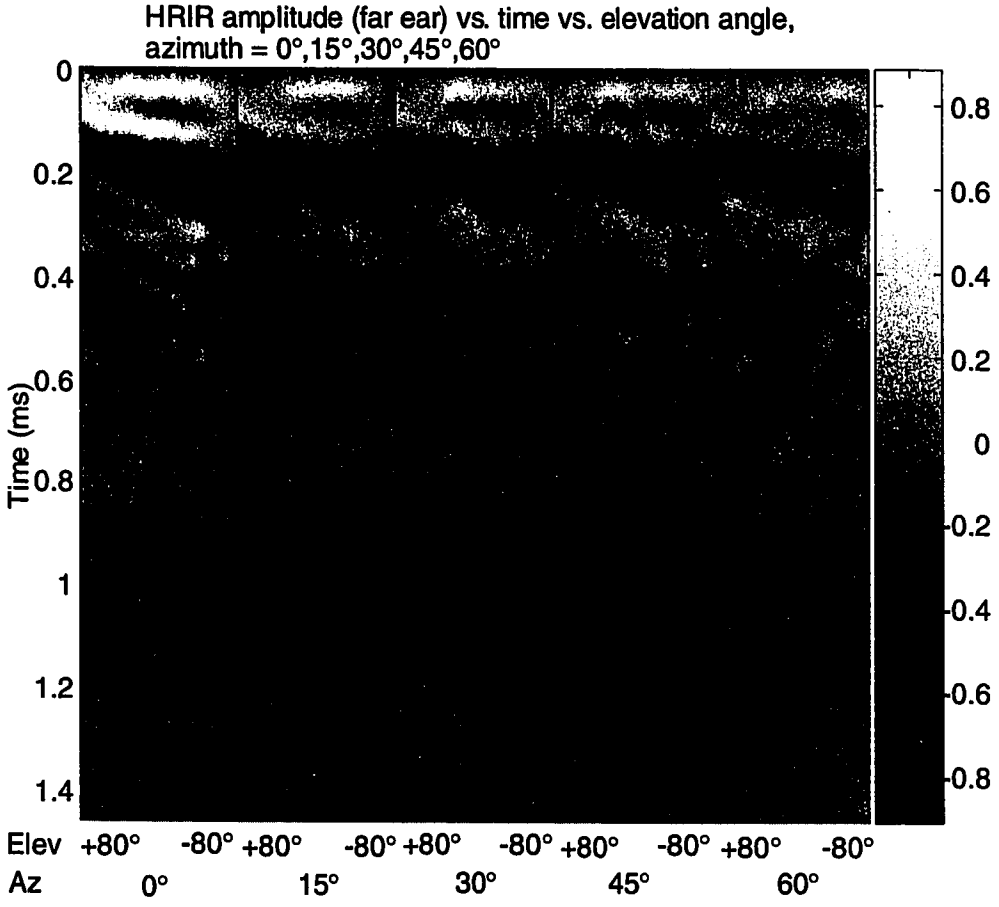


Figure 5a - Composite HRIR Response (Far Ear - Subject PB)

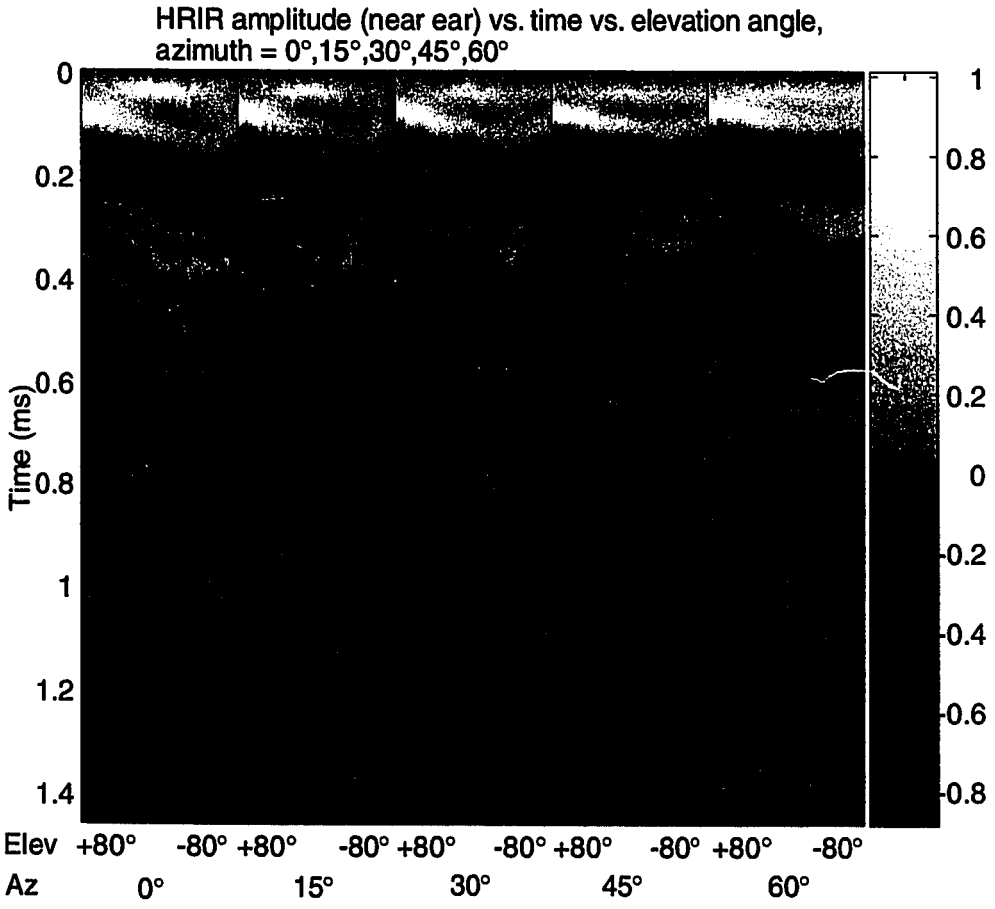


Figure 5b - Composite HRIR Response (Near Ear - Subject PB)

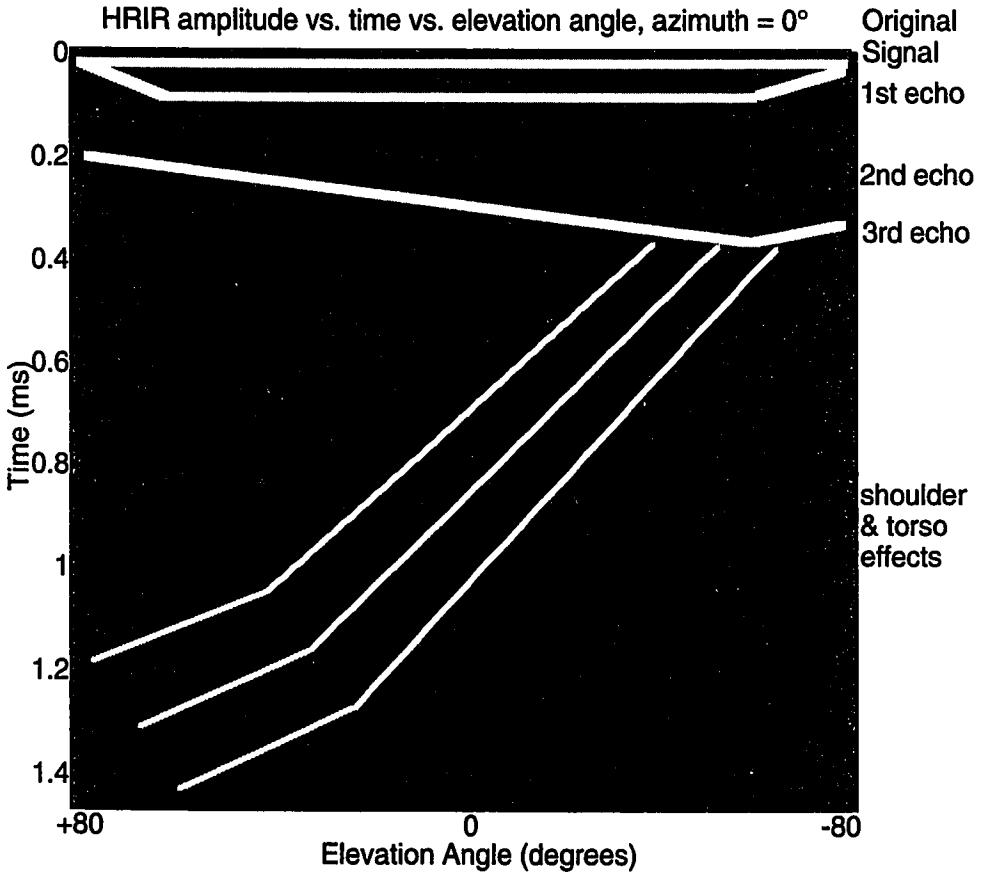


Figure 6 - Gray-scale schematic representing features found in the HRIR responses of Figures 4 and 5

In each gray-scale plot, an initial, non-elevation-varying peak can be seen near the first sample (time t=0 μs). A second peak occurs at t≈50 μs, which varies somewhat with elevation, especially near $\phi=\pm 80°$. A third peak (actually a "valley"), is very prominent. As the sound source is moved from $\phi=+80°$ to $\phi=-80°$, the valley's position in time (with respect to the non-varying initial peak) increases nearly-monotonically from t≈100 μs to 300 μs.

A fourth peak (the third echo) is also apparent, and it follows the same path as the valley (2nd echo), offset by t≈50 μs to 100 μs. Note the similarity of the echoes in the near and far ears now that the head-shadow and timing differences are removed.

This data agrees well with both Batteau [1] and Watkins [12]. The first, non-varying peak is considered to be the initial pulse that reaches the eardrum (or, in this case, the blocked-meatus microphone). The second peak is considered to be an echo whose latency changes both with azimuth and elevation. The third valley is also an echo, and is considered to be the primary monaural cue in detecting elevation [12]. The physical explanation of these echoes is generally attributed to the outer ear (pinna) acting as a reflector. Specifically, the cavum concha is thought to be responsible for causing the elevation-dependent cues [9]. Figure 7 shows a simplistic look at cavum concha reflections at 5 elevations. As the source moves from above to below, the delay of the reflection increases in a similar fashion to the measured data. When the source is very low, the path length begins to shorten again, and this feature can be seen in the data as well. Also, reflections from the cymba concha may play a role when the source is low, as seen in the "split" of the valley echo at lower elevations.

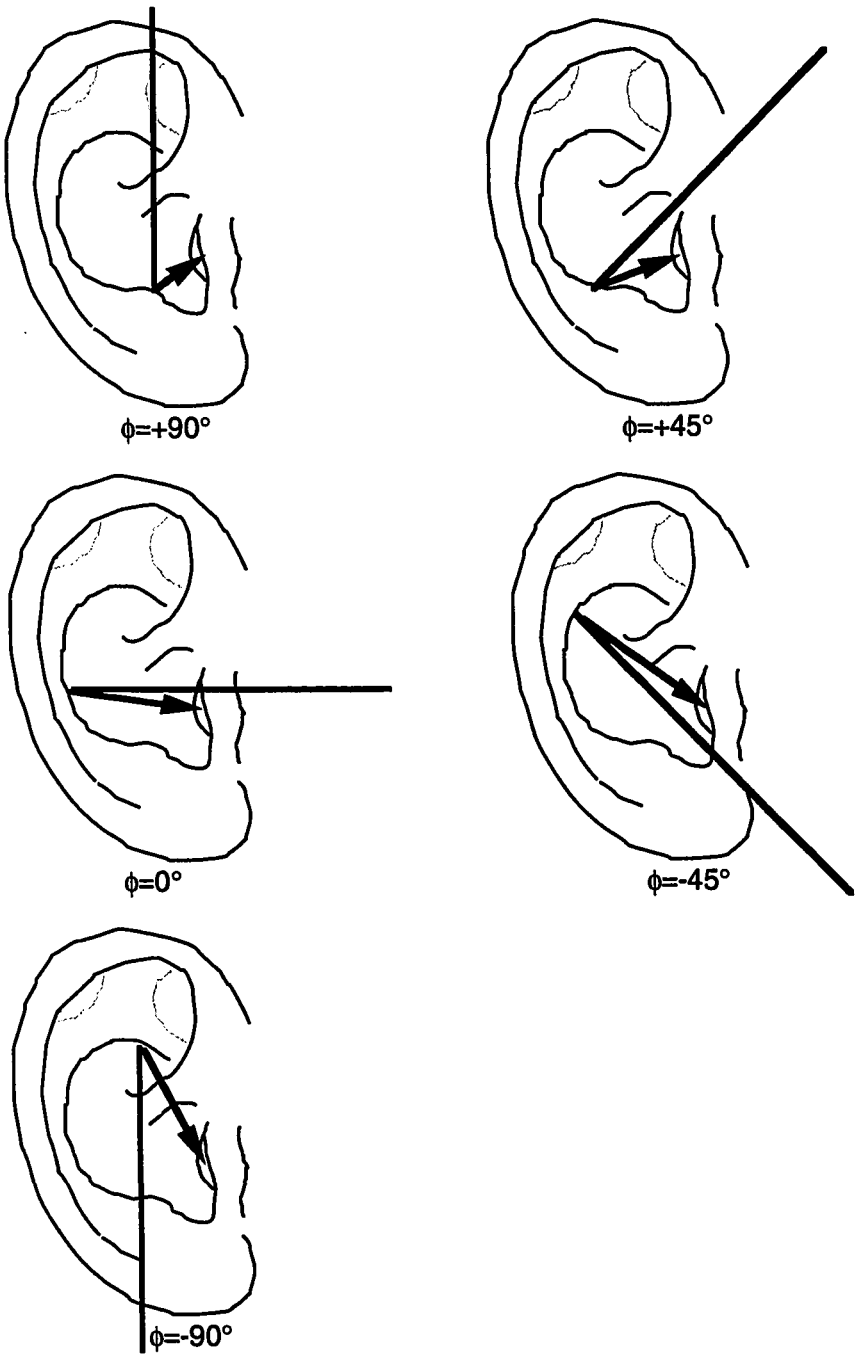


Figure 7 - Pinna echo path length variations based on direct reflection by the cavum concha.

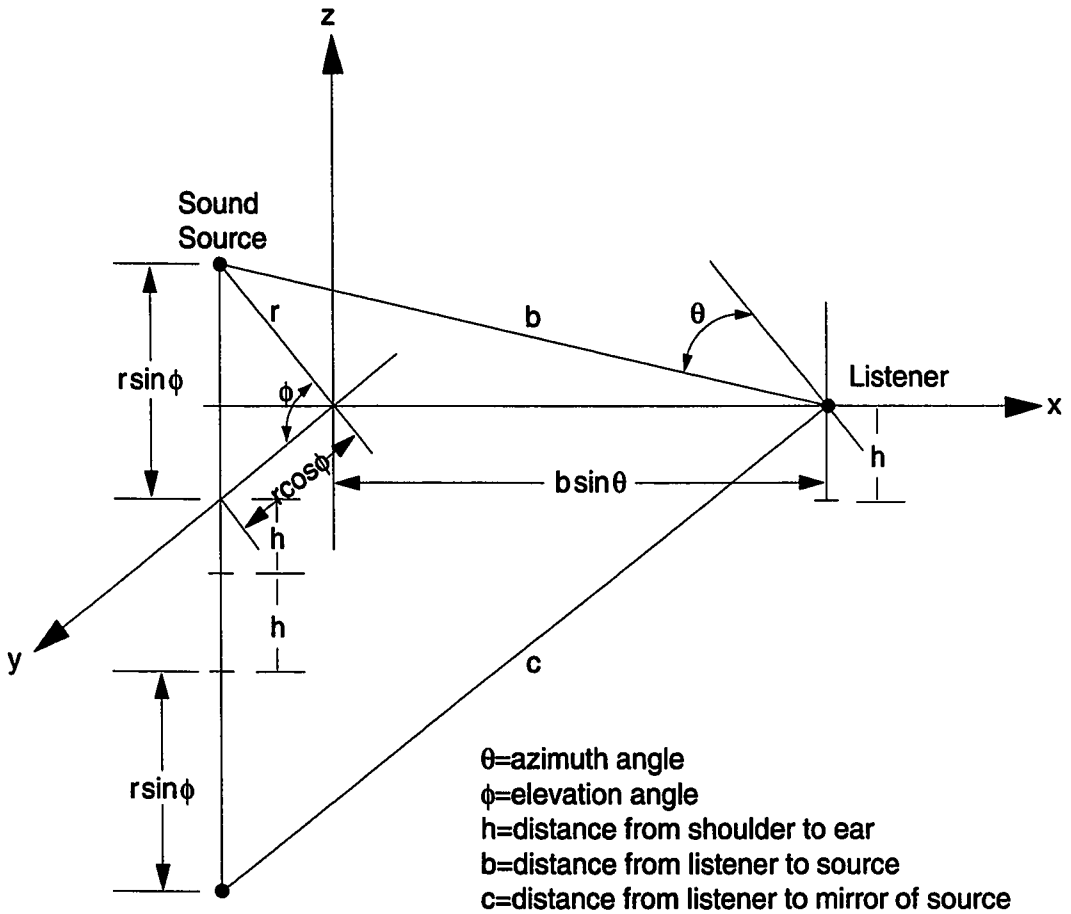
Using HRIR's measured in the median plane, Hiranaka and Yamasaki [10] observed that a "front" source contains two or more major reflections and a "rear" source has only one major reflection. The measured data presented herein also contains at least two major reflections. Hiranaka and Yamasaki also observed that as the sound source approaches "above" (ϕ=+90°), these echoes go away. It can be seen in the measured data that as the sound source approaches ϕ=+80°, the echoes arrive very soon after the initial peak, and merge into a single impulse. The echoes appear to be much more sharp near ϕ=+80° as well.

Another striking detail in the gray-scale plots is the presence of a shoulder/torso reflection. It appears as a series of weak reflections which begin at approximately 1mS from the initial impulse (ϕ=+80°). These reflections grow in intensity and move in closer to the initial impulse as the sound source moves downward. Pinna echoes convert a single shoulder reflection into the series of reflections seen in the data.

From the median plane gray-scale plot, it can be seen that when the sound source was at ϕ=+80° (almost directly above), the weakest shoulder reflection occurred (relative to the other elevations in the median plane). This is most likely due to the shoulder and torso presenting a smaller cross-section when the sound is directly above, as compared to when the source is at lower elevations. When the source is directly above in the median plane, the reflection's extra path length is approximately twice the distance from the ear to the shoulder . For subject PB, the total distance is approximately one foot,

or, in terms of the time it takes sound to travel, approximately one millisecond. This corresponds closely to the delay seen before the first shoulder reflection appears in the gray-scale plot of the median plane.

To better understand this, a simple mathematical model was made of the geometry of the head and shoulders (Figure 8). The model assumed the shoulders were a flat surface and that the sound source was a single plane wave, arriving at an angle in the range $0° \leq \phi \leq +90°$. The reflection appears as a mirrored image of the source, and the delay, angle of arrival, etc. can be determined. The reflection of the wave in the model agreed with the patterns seen in the gray-scale plots. Because the sound arrives at the pinna as a reflection, the apparent elevation angle of arrival of the reflection is different than that for the direct sound. In other words, when the sound source is at ϕ, the shoulder reflection appears to come from approximately $-\phi$. This model is valid only for sound sources in the range $0° < \phi < +90°$. Figure 9 superimposes this shoulder model on top of the actual data.



θ=azimuth angle
ϕ=elevation angle
h=distance from shoulder to ear
b=distance from listener to source
c=distance from listener to mirror of source
r=min. distance from interaural axis to source
x-axis=interaural axis
xy plane= plane of shoulders

Figure 8 - System for shoulder reflection modeling. The mirrored source represents the reflection,. The difference in path length between b and c corresponds to the time delay of the shoulder reflection.

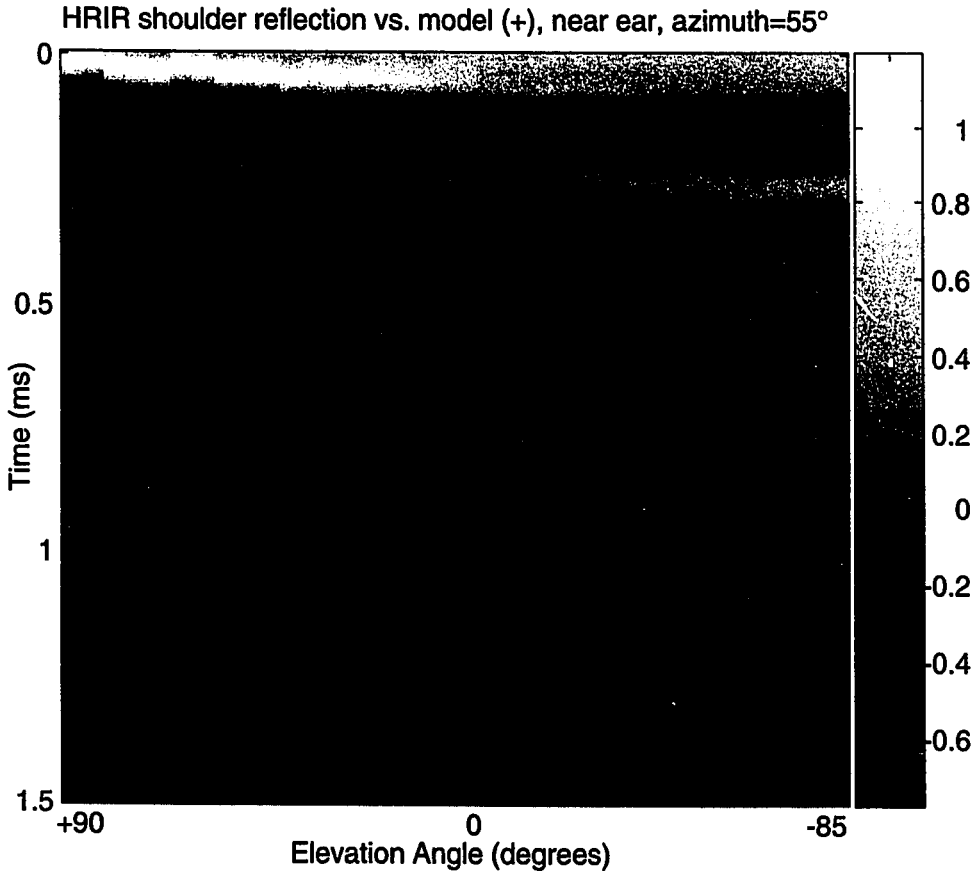


Figure 9 - Measured HRIR shoulder reflection
vs. physical model (near ear - subject PB)

When the sound source is below, the torso provides a greater surface area for sound to scatter off, which is seen in the gray-scale plots as a increase in the amplitude of the echoes. Also, because the sound is scattered earlier (due to the proximity of the source to the torso), it arrives at nearly the same time as the direct sound. This, too, is clearly seen in the gray-scale plots.

The Model

Early in the modeling process it was decided to base the work in the time-domain. We consider the detailed timing information presented in the HRIR to be critical in developing an accurate model. A modeling approach in the frequency-domain may yield an apparently close match in the magnitude response, but reproducing the phase response is difficult.

In developing a model of the elevation characteristics of the HRIR, it was decided to approach the task from a structural perspective. Genuit's model (Figure 2) takes a structural approach, but treats the head, shoulders, pinna and upper body as a sum of contributors to the cavum concha input. We chose a model with a combination parallel/serial approach (Figure 3), with the components arranged for maximum simplicity. Physically, the signal should be passed through the head-shadow and ITD blocks before reaching the pinna blocks. However, since the model is linear, we can model the pinna and shoulder echoes first, with the head-shadow and ITD following. This allowed for a more compact, although somewhat less intuitive, model.

The pinna echoes are each modeled as a delay stage, along with a scaling factor or gain. The same method is also used for the shoulder echo. These delayed signals are summed with the original, and then passed through a "smoothing" filter, which accounts for the duration or "spread" of the echo. This spread was determined by weighted averaging of the impulse responses. The signal is then split into right and left channels, at which time the

corresponding head-shadow and ITD are introduced. The head-shadow filter is simply the inverse of the "inverse-head-shadow" filter which was used earlier on the raw HRIR data. Therefore, apart from the head-shadowing and timing differences (azimuth-dependent cues), the same pinna-echo model is applied to both ears.

The decision to use the same pinna-echo model for both ears was arrived at for several reasons. The first is that, upon examining the enhanced data, only slight differences in timing and amplitude were observed between the near-ear and far-ear HRIR responses. Secondly, the addition of a second channel takes away from the simplicity of the model. Thirdly, if a second channel were to be added, it is unclear what features significantly distinguish it from the first channel.

While the second echo (the large valley in the plots) moves nearly-monotonically in time with respect to the initial peak, it does not move linearly. It also has a slight dependence on azimuth angle. Near "below" (ϕ=-80°) this valley splits into two parts, as if another echo has begun to interfere with it. This feature is more pronounced near the median plane and diminishes as the source moves toward θ=60°. No effort was made to reproduce this feature due it's complexity. The valley's latency has a somewhat sinusoidal change with respect to elevation. The rate of change is constant as the sound source moves downward from above, until about ϕ=-45°, when it flattens out and begins reversing direction.

Equation (1) describes the amount of delay for the model's pinna echoes (τ_p):

$$\tau_{pn}(\theta,\phi) = A_n \cos(C_n\theta)\sin\left(D_n\left(\phi + \pi/2\right)\right) + B_n \quad \text{for } 0 \le \theta \le \pi/2 \ , \ -\pi/2 \le \phi \le \pi/2 \tag{1}$$

where *A* is a slope, *B* is an offset, *C* and *D* are scaling factors and *n* indexes the *nth* echo.

The steepness of the change is controlled by the slope (A), while the initial delay is controlled by the offset (B). The factors C and D allow for some minor scaling, and turn out to have values close to 0.5 and 1, respectively. The measured data was used to establish initial values for the coefficients A, B, C and D. Using both an empirical "trial and error" approach and system identification techniques, a set of values for the coefficients was arrived at. The first pinna echo has a unique set of A, B, C and D coefficients, since it is more closely related to azimuth than elevation [1,11]. The remaining 2 through *n* pinna echoes are closely related to each other and use the same A, C, and D coefficients; the offset B increases as a factor of n.

Equation (2) describes the delay of the shoulder/torso reflection (τ_{sh}) :

$$\tau_{sh}(\theta,\phi) = As\sin(Cs\theta)\cos\left(Ds\left(-\phi + \pi/2\right)\right) + Bs \quad \text{for } 0 \le \theta \le \pi/2 \ , \ -\pi/2 \le \phi \le \pi/2 \tag{2}$$

where *As* is a slope, *Bs* is an offset, *Cs* and *Ds* are scaling factors.

As with the pinna delay, the steepness of the change is controlled by the slope (A), while the initial delay is controlled by the offset (B). The factors C and D again allow for minor scaling adjustments. Note that this model is fitted to the measured data, while the model used in Figure 9 was derived using body geometry.

We originally allowed the magnitude of the pinna echoes (ρ_{pn}) to be functions of azimuth and elevation. These were later simplified to constant scaling factors, as listening tests indicated that the magnitude was not an important cue in detecting elevation. The sum of the magnitudes of the scaling factors adds up to 1 at each elevation, providing a dc gain of 0dB. This provides a flat frequency response at the lower frequencies which do not contribute to elevation cues.

Alone, this FIR model provides important elevation cues when used to filter broad-band noise. The frequency response of this filter provides spectral notches similar to the notches seen in the measured HRIR data. The FIR filter frequency response does not, however, match the overall spectrum of the measured data. The measured data shows roll off at the high end of the spectrum, so we decided to include a low-pass filter in the model to help approximate the measured data's roll-off. The time-domain response was equally improved using this filter, as seen below in the performance section. Several low-pass filter representations were tried, based on weighted averaging of the impulse responses. A simple first-order Butterworth low-

pass filter was eventually decided upon, as it approximated the high-frequency roll-off fairly well.

Originally, ear-canal resonance was going to be accounted for in the model. The blocked-meatus microphones that were used to make the measurements occupy the outer portion of the ear-canal, and should prevent any ear-canal resonance from appearing in the measured data. The model, which is based on the measured data, also omits the ear-canal resonance. During play back, the listener's own ear-canal resonance comes into play (due to the use of headphones or in-ear phones). If the model did account for ear-canal resonance, the listener would be subjected to hearing the resonance twice: his/her own and the model's.

The overall model functions as follows: the delay (τ_p) and magnitude (ρ_{pn}) coefficients are calculated and used to create a finite impulse response (FIR) filter that is no more than 32 samples long (32 samples is approximately 725 μS, which allows enough time for the pinna echoes to appear). A monaural source is input and filtered using this FIR filter, and then low-passed using the first-order Butterworth filter. The response is then split and filtered into left and right channels using the head-shadow and ITD models (Appendix A), which provide a monaural to binaural transformation.

Figure 10a shows the measured HRIR response (head-shadow removed) for the near ear at an azimuth angle θ=55° for subject PB. Figure 10b is the modeled HRIR response (including the pinna and shoulder echoes), without

the smoothing filter. Figure 10c is the model with the smoothing filter. Table 1 provides a summary of the coefficients used for each subject. Appendix D contains the Matlab code used to synthesize the model.

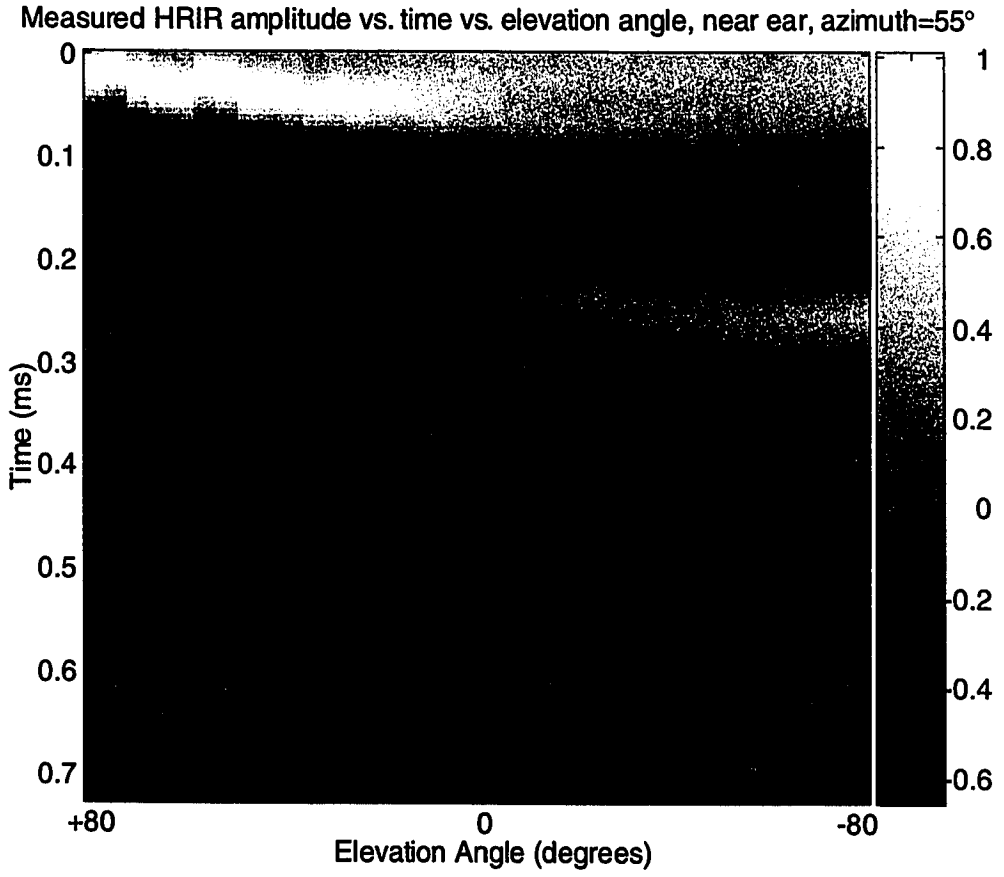


Figure 10a - Measured HRIR (head-shadow removed) for subject PB (θ=55°)

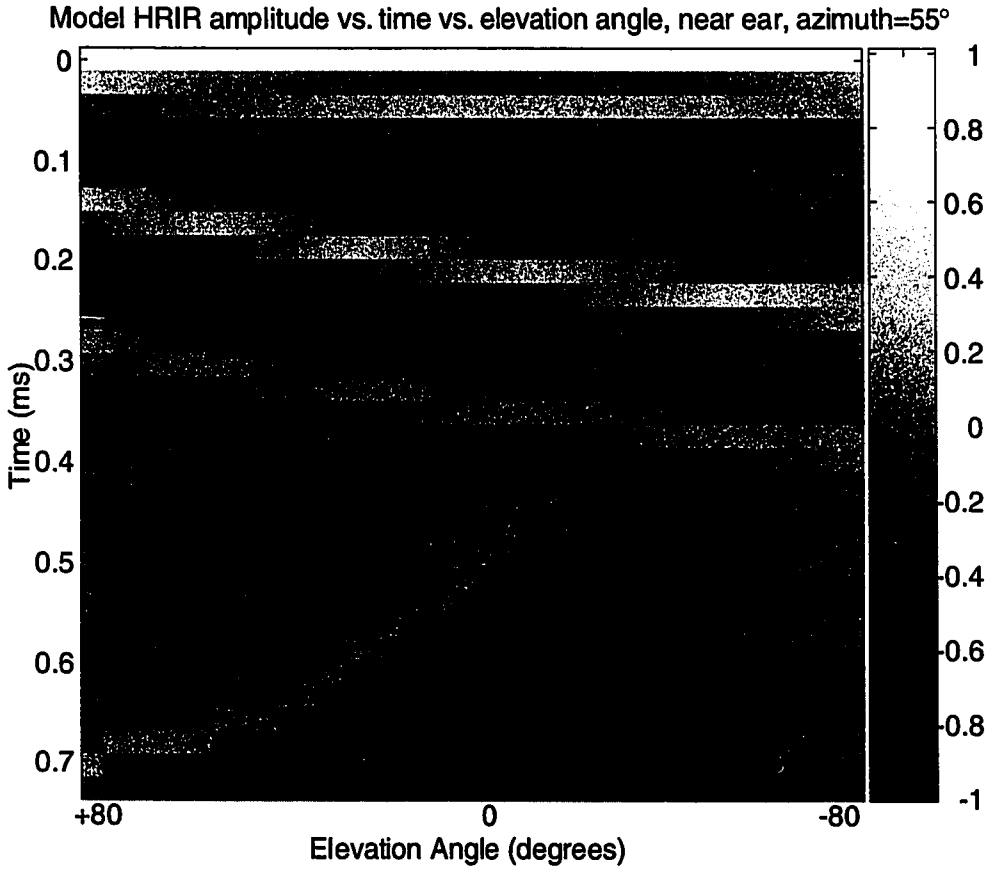


Figure 10b - Model: pinna and shoulder echoes, no smoothing filter, θ=55°

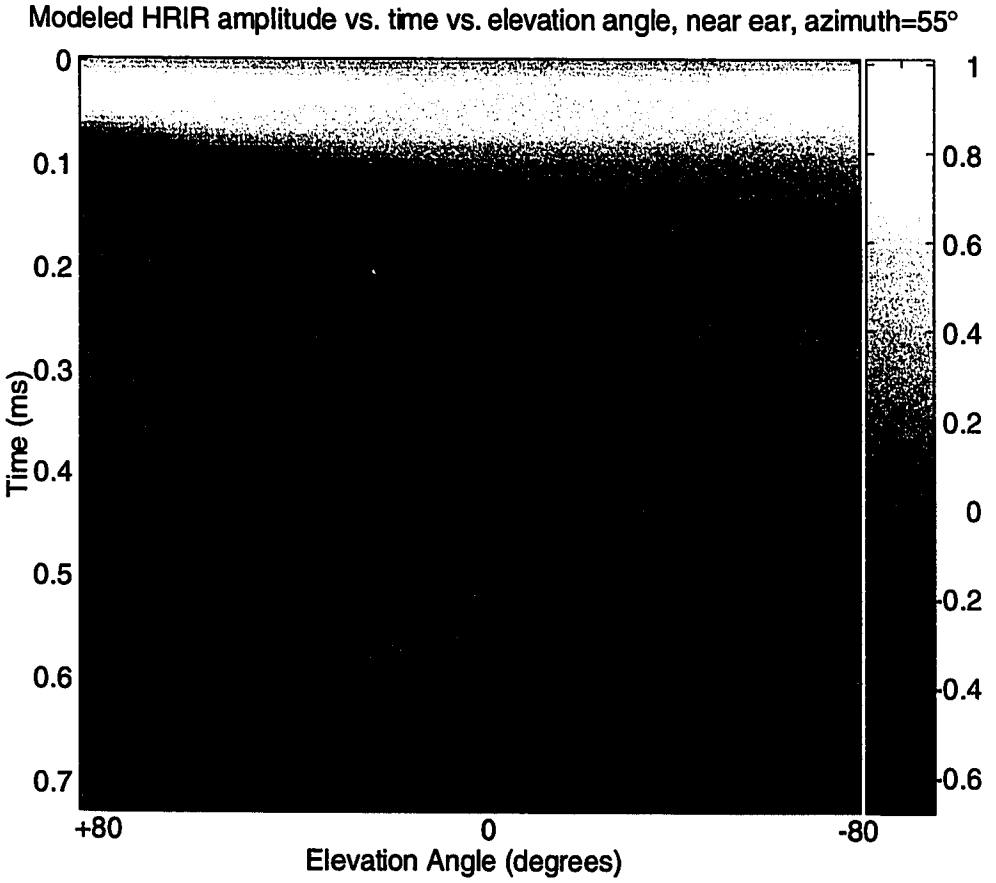


Figure 10c - Model: pinna and shoulder echoes, with smoothing filter, θ=55°

Coeff	PB	NH	RD
A1	1	1	1
A2	5	5	5
A3	5	5	5
A4	5	5	5
A5	5	5	5
B1	2	2	2
B2	4	4	4
B3	7	7	7
B4	11	11	11
B5	13	13	13
C1	0.5	0.5	0.5
C2	0.5	0.5	0.5
D1	1	1	0.85
D2	0.5	0.5	0.35
As	44	44	44
Bs	12	12	12
Cs	0.5	0.5	0.5
Ds	0.6	0.6	0.6

Table 1 - Coefficients used in Model

Listening Tests

Listening tests were performed to evaluate the effectiveness of the model. The three individuals who participated in the original measurements at $\theta=55°$ and $+90° \leq \phi \leq -85°$ were the subjects. All tests were implemented in identical fashion: the subjects listened to a 500-mS burst of Gaussian white noise that was played over headphones (or in-ear phones) via a computer. The headphones used were Sony model MDR-31 and the in-ear phones were Etymotic model ER-2.

A graphical user interface was provided that allowed the subject to select a noise to be played from a random elevation (Figure 11). In each case, the randomly selected noise had been first convolved with the subject's measured HRIR. The subject was then asked to ignore any timbre differences and to match the perceived elevation of the randomly selected noise burst to one of a set of 35 noise bursts that could be selected with a slider corresponding to elevations from +90° to -85°. These noise bursts originated from versions of the model that had been convolved with noise, or, in the case of the baseline tests, the same measured HRIR data as used for the randomly selected noise burst. No restrictions were placed on the number of times the subject could listen to either the original random noise or the bank of noises before making the final choice for the best match.

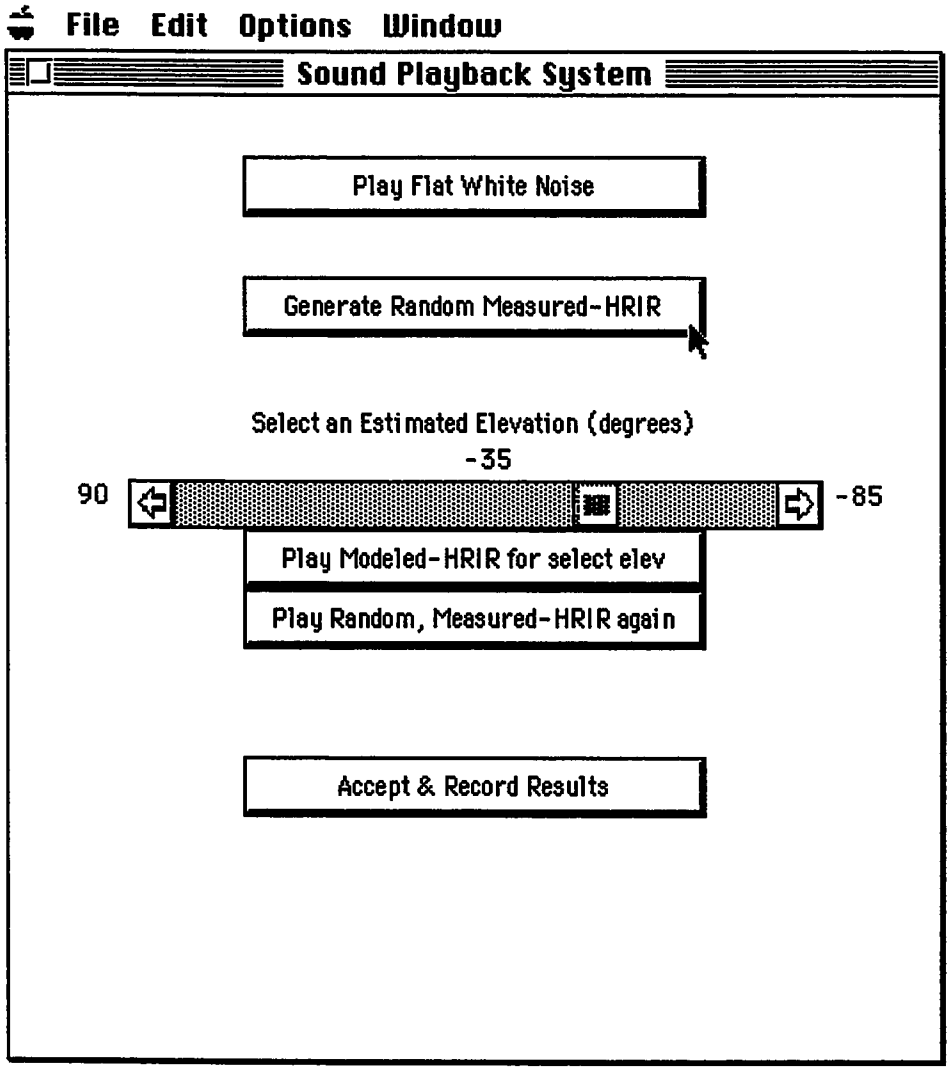


Figure 11 - Graphical Interface for Sound Playback

To establish a baseline, the subjects were asked to listen to a Gaussian white noise burst that had been convolved with the subject's randomly selected HRIR (random elevation). The subject was then asked to match it to a bank of noise bursts which corresponded to the subject's measured HRIR at each

elevation. The results for the three subjects are shown in Figures 12a,b,c (headphones) and Figures 13a,b,c (in-ear phones). Several trends are observable from this data. The first trend was an overall tendency to place the sound somewhat "higher" than the corresponding elevation. The second trend was a reduced accuracy near the elevation end-points (-85° and +90°). The subjects perceived the sound as "higher" when it was near -85° and "lower" near +90°. The error was slightly worse for all subjects using the headphones, so the remaining tests were performed only with the in-ear phones.

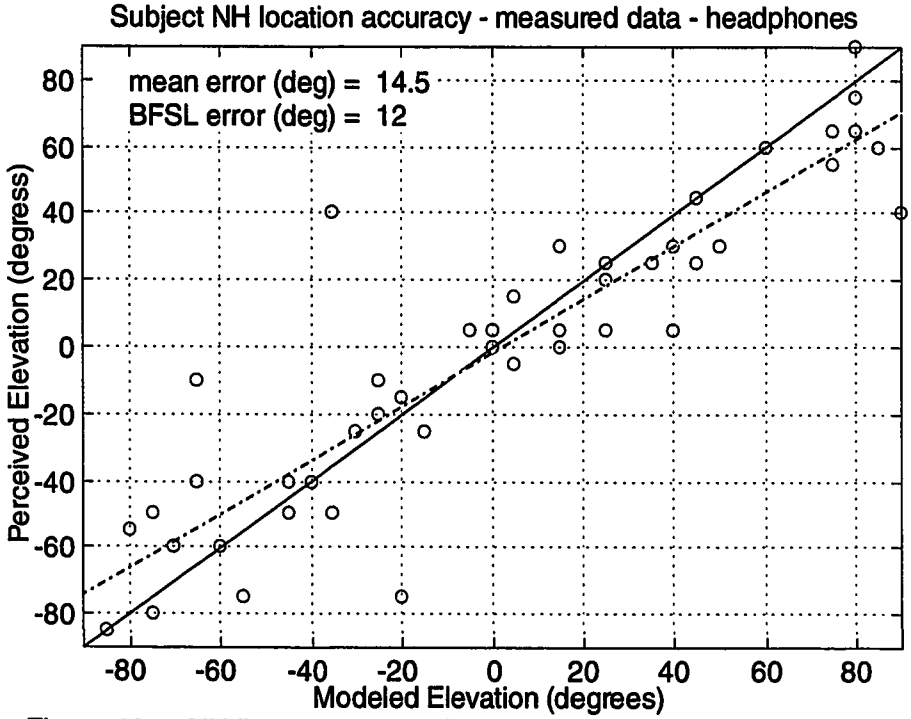


Figure 12a - NH listening test performance (measured vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

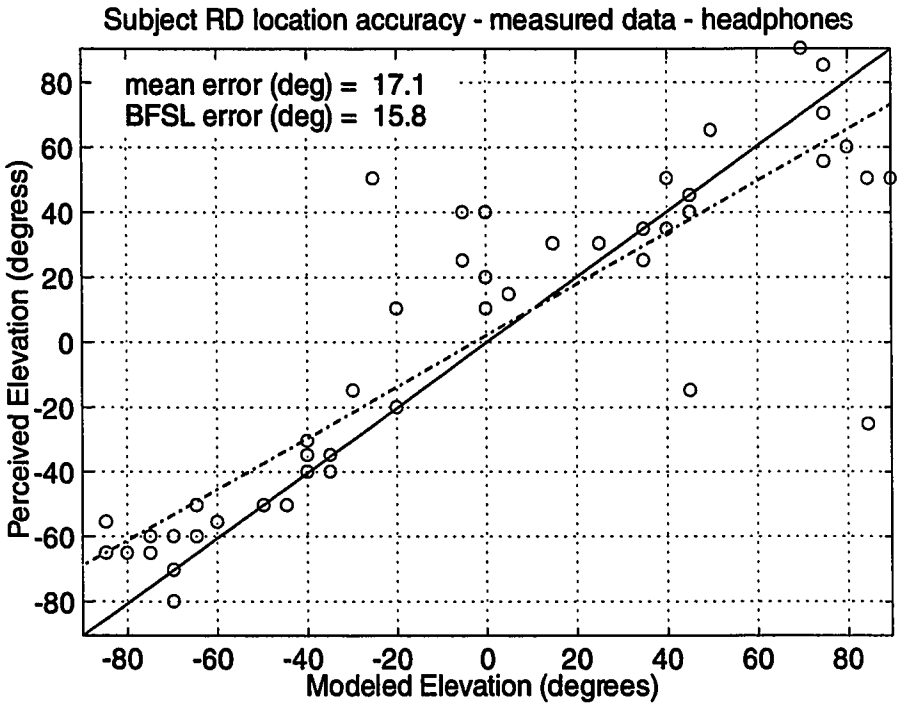


Figure 12b - RD listening test performance (measured vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

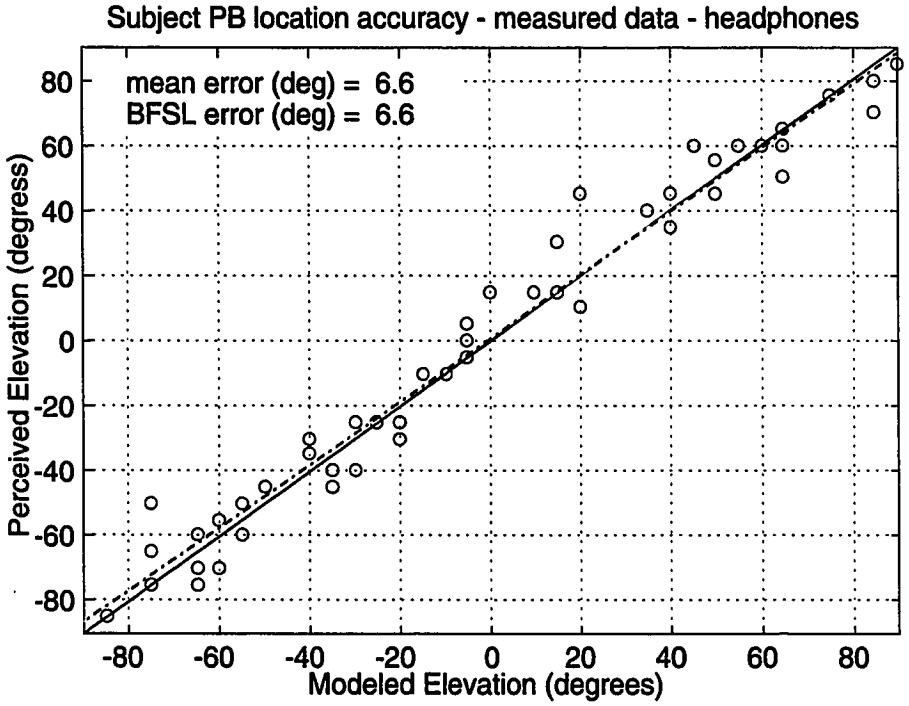


Figure 12c - PB listening test performance (measured vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

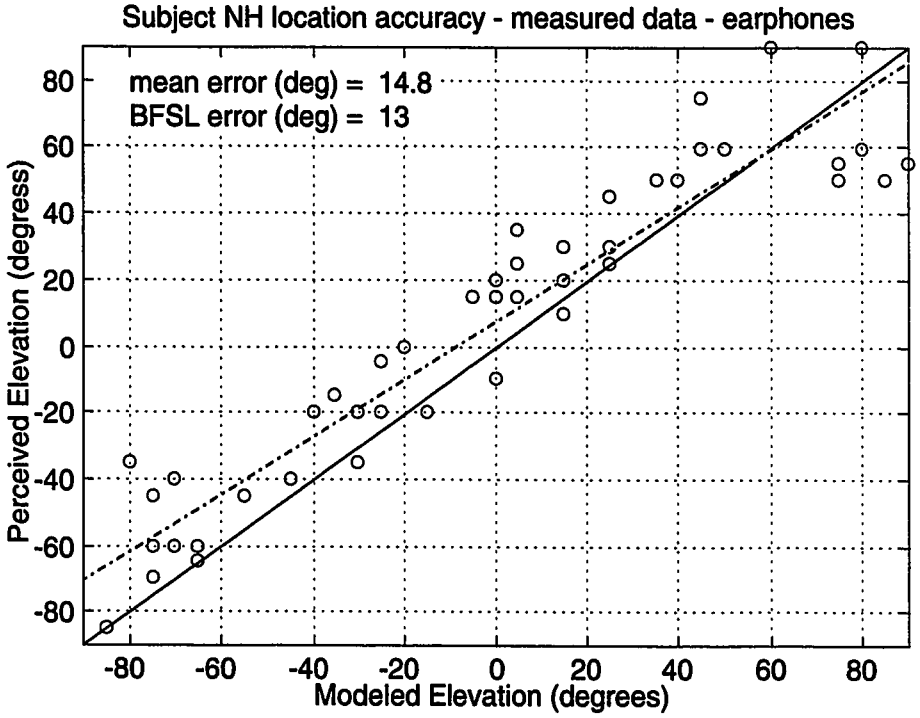


Figure 13a - NH listening test performance (measured vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

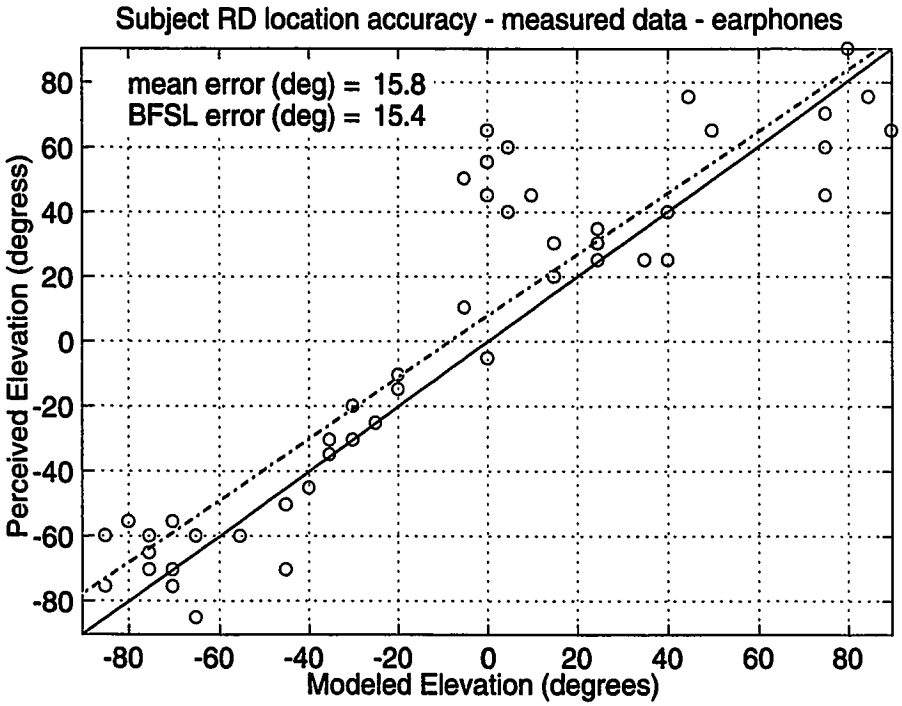


Figure 13b- RD listening test performance (measured vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

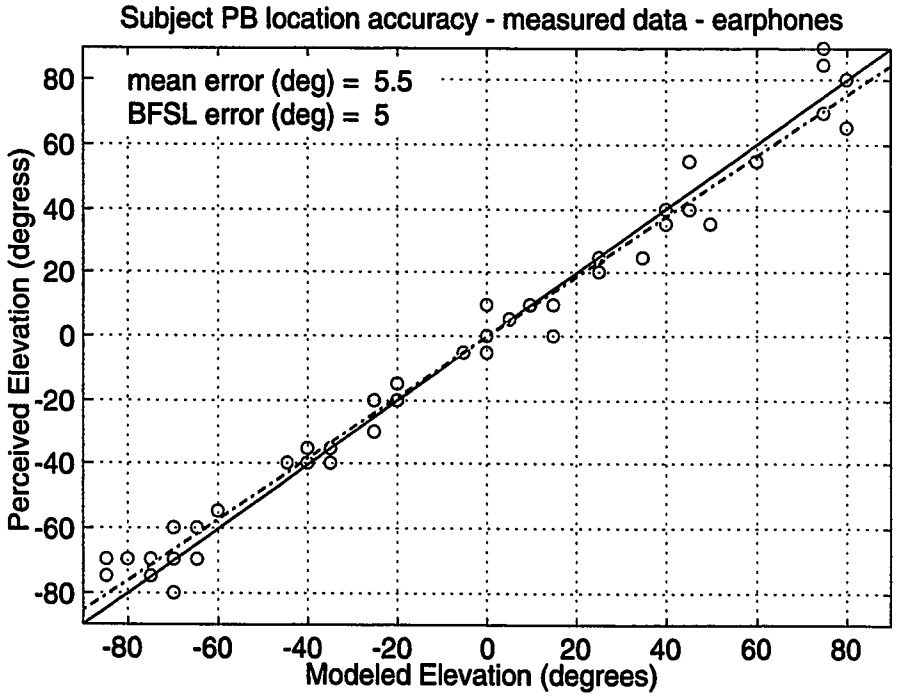


Figure 13c- PB listening test performance (measured vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

To validate the model, the same subjects were asked to listen to a Gaussian white noise burst that had been convolved with the subject's randomly selected HRIR (random elevation), and then to match it to a bank of noise bursts that corresponded to the modeled HRIR for each elevation convolved with noise. The results for the three subjects are shown in Figures 14a,b,c. Again, a reduced accuracy near the extreme elevations (-85° and +90°) was apparent. Additionally, the overall accuracy was worse than the previous test (which used the subject's measured HRIR for the matching bank). Subject RD has a tendency to place model higher than the corresponding elevation, but

this trend was not as apparent in the other two subjects.

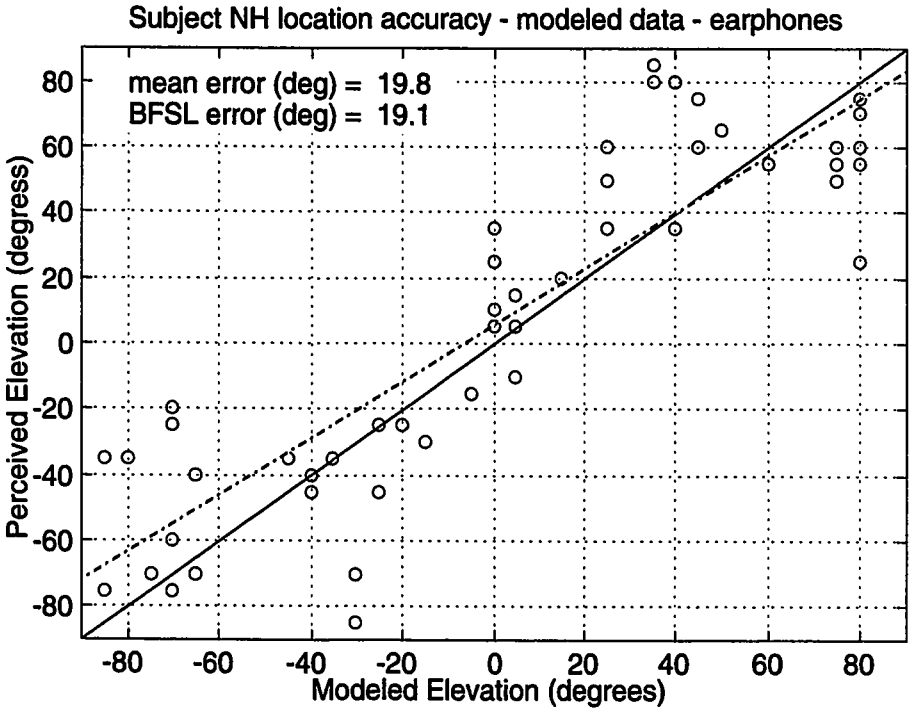


Figure 14a - NH listening test performance (model vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

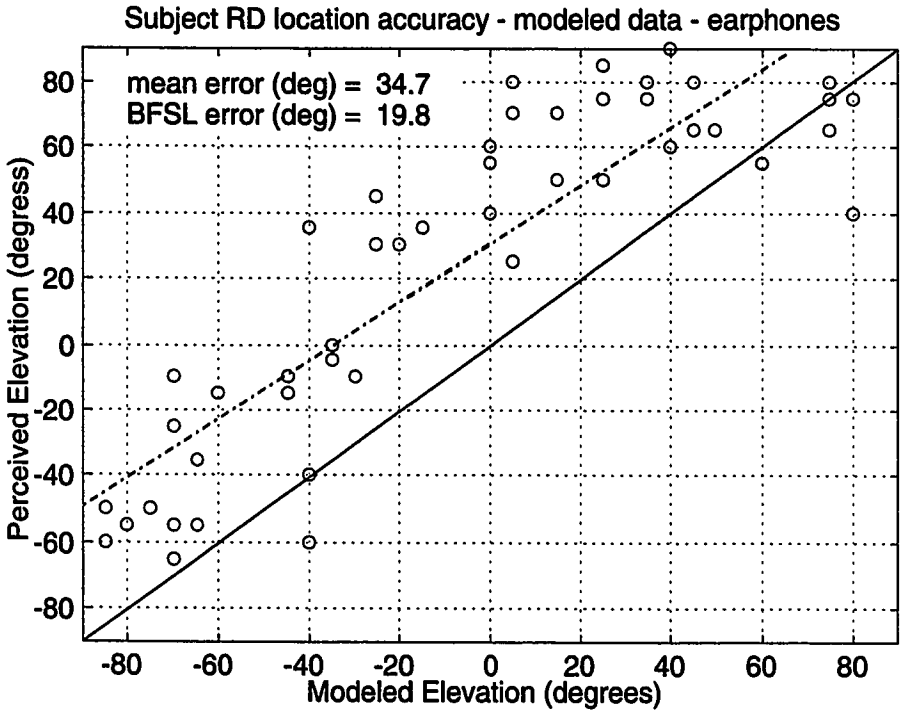


Figure 14b - RD listening test performance (model vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

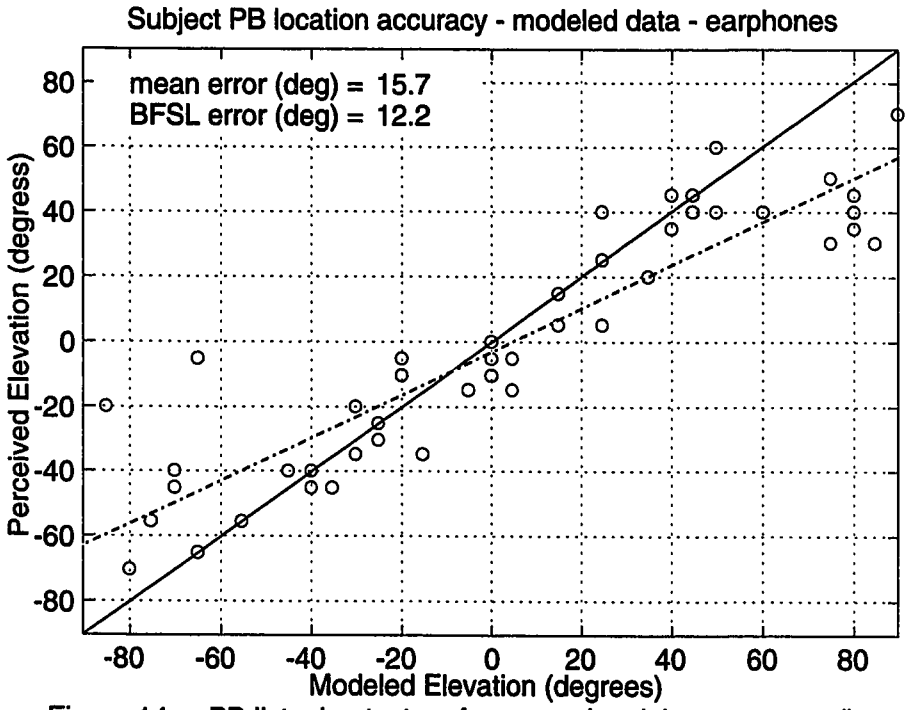


Figure 14c - PB listening test performance (model vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

In addition to validating the model described in this thesis, it was decided to contrast the model with the pinna-echo model Watkins [12] used for his psycho acoustic tests. Watkins' model was based on Batteau's pinna-echo work. Watkins summed a direct (no delay) signal with both a fixed delay and a variable delay "echo" and then convolved them with noise. The fixed delay was 39μs, while the variable delay ranged from 97μs to 312μs. The magnitude of the delays was set to unity (as was the direct signal), and the polarity of the delays could be inverted. Watkins showed that moving the variable delay resulted in an elevation cue perceived by the test subjects.

To contrast the performance of a "delay-and-add" system such as Watkins' with the model described herein, we performed another set of matching tests. The same subjects were asked to listen to a Gaussian white noise burst that had been convolved with the subject's randomly selected HRIR (random elevation), and then match it to a bank of noise bursts (which corresponded to the "delay-and-add" system convolved with noise). The "delay-and-add" bank used a fixed delay of 45 μs (a 2 sample delay) and a variable delay of 97μs to 312μs. The gains were set to unity, and the polarity of the variable delay was inverted (Figure 15). No other processing was done to the signal. The "delay-and-add" noise bursts appeared only in the near (right) ear of the test subjects (monaural presentation). The results of the tests for the three subjects are shown in Figures 16a,b,c.

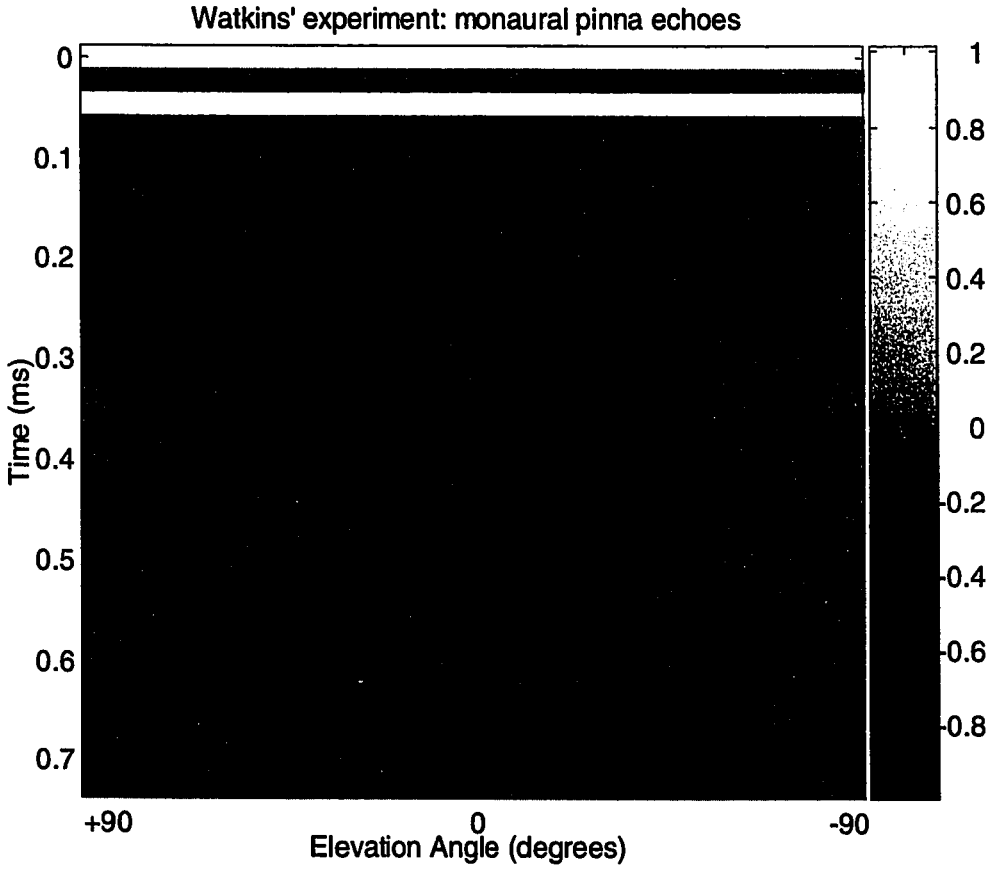


Figure 15 - Watkins' experiment: linear, monotonic delay

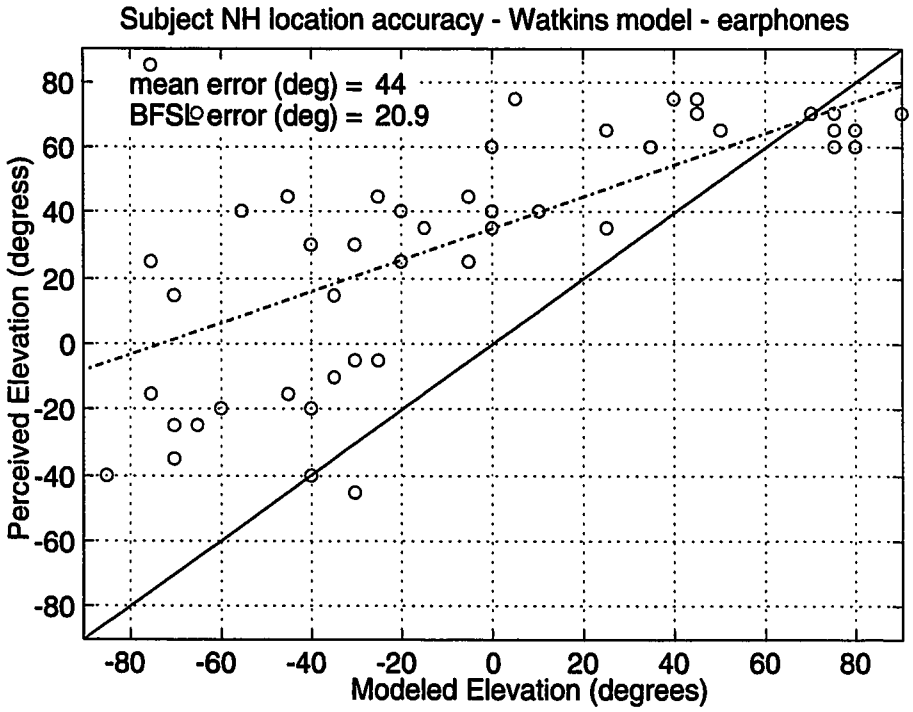


Figure 16a - NH listening test performance (monaural vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

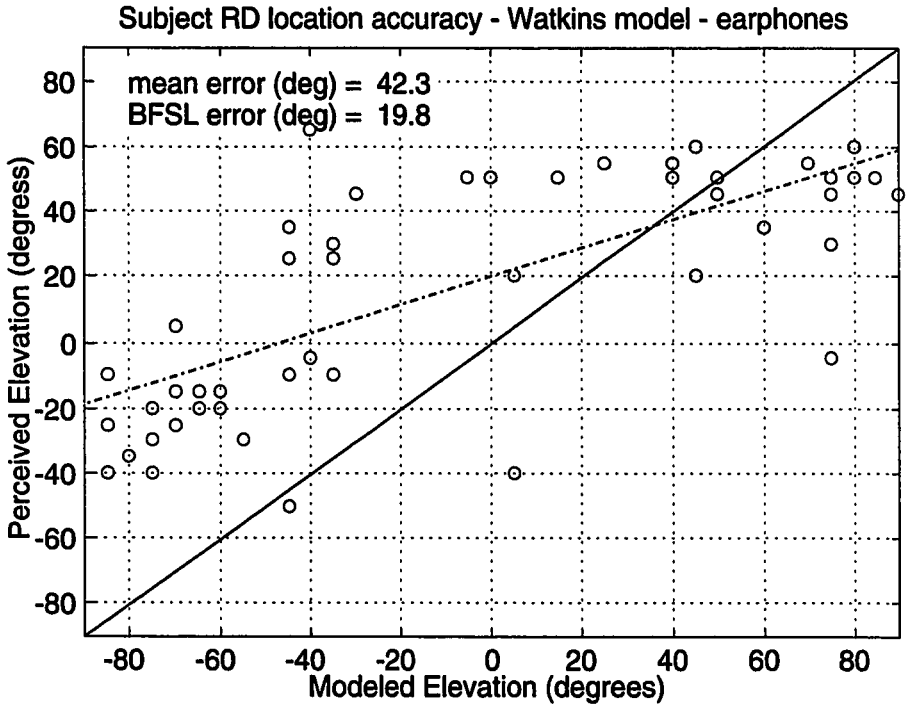


Figure 16b - RD listening test performance (monaural vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

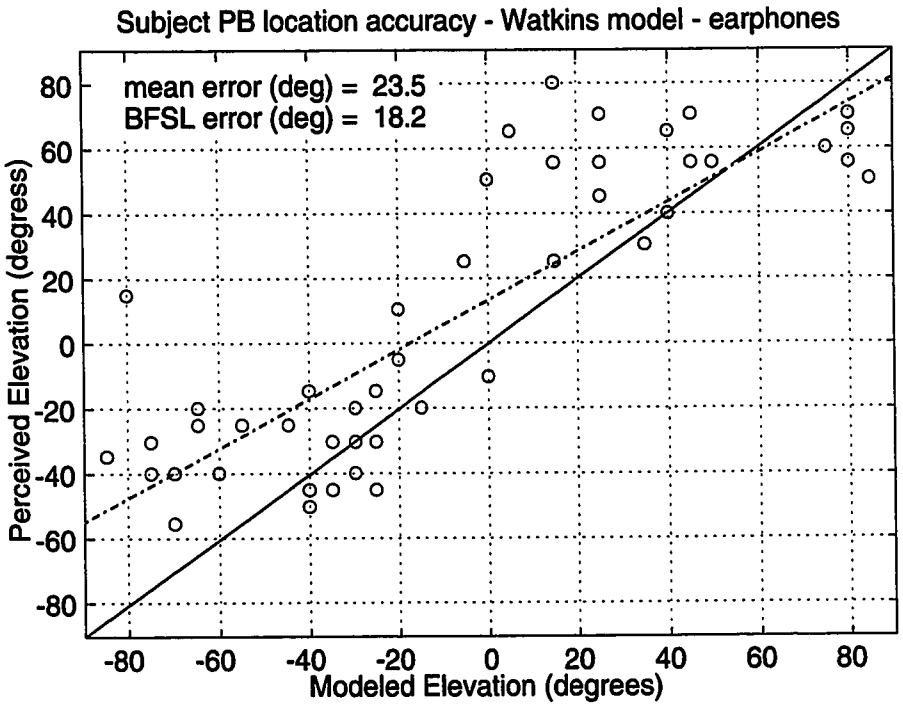


Figure 16c - PB listening test performance (monaural vs. measured)
o = data point
broken line = best fit straight line fit to data
solid line = ideal match

The monaural "delay-and-add" model provided elevation cues, as seen in the test data trends. It was, however, significantly worse at generating correctly perceived elevation cues in comparison to the model presented within this thesis. These results indicate that the monaural cue is important in detecting elevation, but a properly synthesized binaural cue, adapted to the individual listener, significantly improves the accuracy.

One artifact noticed by the subjects was that the monaural "delay-and-add" system had a "reversal" in the perceived elevation. Once the sound was

outside of the range $-45° \leq \phi \leq +45°$, the sound appeared to reverse direction or "fold in" upon itself. This reversal agrees with Watkins' observations. The thesis model did not appear to suffer from this type of reversal, but it was limited in the perceived range of elevation.

Table 2 provides a summary of the mean errors that resulted from the listening tests. The columns are arranged by test subject. The rows are divided between the four sets of tests. Each test is then split into two additional rows. The first of those two rows corresponds to the mean error between the perceived location and the actual (ideal) location. The second row is the mean error between the perceived location and the best-fit straight line corresponding to the perceived data.

		NH	RD	PB
Measured (headphones)	mean error	14.5°	17.1°	6.6°
	bfsl error	12.0°	15.8°	6.6°
Measured (earphones)	mean error	14.8°	15.8°	5.5°
	bfsl error	13.0°	15.4°	5.0°
Model (earphones)	mean error	19.8°	34.7°	15.7°
	bfsl error	19.1°	19.8°	12.2
Watkins (earphones)	mean error	44.0°	42.3°	23.5°
	bfsl error	20.9°	19.8°	18.2°

Table 2 - Mean errors for listening tests

It is interesting to note that the monaural "delay-and-add" model scores very

closely to the binaural model with respect to the best-fit straight line data. The two models differ greatly, however, with respect to absolute error (deviation for the ideal), as the binaural model is clearly superior to the monaural "delay-and-add" model.

Conclusions

The HRIR measurements presented herein agree well with Batteau [1]. Upon removing the effects of head-shadowing and ITD, the measured HRIR data clearly shows a valley that moves with respect to elevation for all three subjects. This feature is present in both the near and far ears. Additional elevation dependent features (such as the shoulder echo) are also apparent in the measured data.

It has been shown that gross elevation cues can be synthesized from a very simple monaural pinna echo model, as in the "delay-and-add" system used by Watkins [12]. The binaural model presented herein, by providing the proper echo latency, head-shadow and ITD information, significantly improves localization over the monaural model.

Both the measured HRIR data and the listening tests indicate that, while pinna echoes provide a strong elevation cue, humans require a more complicated system than a simple pinna-echo model.

Lastly, limitations in the perceived range of elevation exist in the measured

HRIR data, the Watkins "delay-and-add" system and the model presented herein. This may be attributable to headphone and in-ear phone reproduction, since all of the listening tests indicated this limitation to some degree.

Areas for further investigation

There was a slight perceptual bias in the subjects to place the perceived sound as slightly higher than it was measured. Initially, we thought this might be due to conflicting pinna cues when using headphones. The measured data contains pinna cues for specific elevations, but once reproduced over headphones, a second set of pinna cues could be introduced that indicate an azimuth of θ=90° and an elevation of ϕ=0°. The in-ear phones eliminate this possibility, and while all three subjects felt the in-ear phones provided a more convincing presentation, the data reflects only a slight improvement over headphones.

One aspect of modeling elevation cues has to do with the effects of localization versus externalization. Localization is closely related to the HRIR response due to ear, head and shoulder geometry. Externalization tends to be connected to room ambiance, rather than body geometry. Therefore, a perceptually convincing elevation model should incorporate both HRIR modeling and room modeling. The model should be expanded to take into account room reverberation.

During the initial processing of the measured data, it became clear that, while generally accepted as sufficient for audio recording, a sampling rate of 44.1 kHz was inadequate in visually perceiving the detailed structure of the HRIR. Interpolation by a factor of four remedied this problem, but further measurements using a much higher sampling rate (such as 88.2 kHz) are warranted. Additionally, a wide-bandwidth sound source, such as the acoustic impulse generated by a spark, would be useful in eliminating the limited bandwidth of a conventional loudspeaker.

References

[1] Batteau, "The role of the pinna in human localization," *Proc. Royal Society London*, vol. 168 (series B), pp. 158-180 (1967)

[2] Blauret, J. P., *Spatial Hearing*, Cambridge, MA, MIT Press (1983)

[3] Cassaro, Tom M. and Mark J. Van Belleghem, "Implementing time-variable DSP filters to synthesize binaural sounds," Dept. of Elect. Engr., San Jose State University, Technical Report No. 2, NSF Grant No. IRI-9214233 (May, 1993)

[4] Duda, Richard O., "Modeling head related transfer functions," *Proc. Twenty-Seventh Annual Asilomar Conference on Signals, Systems and Computers* (Asilomar, CA, November, 1993)

[5] Duda, Richard O., "Estimating Azimuth and Elevation from the Interaural Intensity Difference," Technical Report No. 4, NSF Grant No. IRI-9214233, Dept. of Elec. Engr., San Jose State Univ., (September, 1993).

[6] Foster, S. H. and E. M. Wenzel, "Virtual acoustic environments: The Convolvotron," in *SIGGRAPH 91 (18th ACM Conference on Computer graphics and Interactive Techniques*, Las Vegas, NV, 1991)

[7] Genuit, Klaus, "A model for the description of outer-ear transmission characteristics," Ph.D. dissertation, Rheinisch-Westfalischen Technichen Hochschule Aachen, Aachen, Germany (December 21, 1984)

[8] Genuit, Klaus and Wade R. Bray, "The Aachen head system," *Audio*, pp. 58-66, (December, 1989)

[9] Hebrank, Jack and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc Am.*, vol. 56, pp. 1829-1834 (1974)

[10] Hiranaka, Y. and H. Yamasaki, "Envelope representation of pinna impulse responses relating to three-dimensional localization of sound sources," *J. Acoust. Soc Am.*, vol. 73, pp. 291-296 (1983)

[11] Kistler, D. J. and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum phase reconstruction," *J. Acoust. Soc. Am.*, vol. 91, pp. 1637-1647 (March 1992)

[12] Watkins, A. J., "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Am.*, vol. 63, pp. 1152-1165 (April 1978)

[13] Wenzel, E. M. et al., "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, pp. 111-123 (July 1993)

[14] Wright, D., J. H. Hebrank and B. Wilson, "Pinna reflections as cues for localization," *J. Acoust. Soc. Am.*, vol. 56, pp. 957-962 (September 1974)

Appendix A - Head-Shadow and ITD Models

The transfer function of Lyon's head-shadow model is of the form:

$$H_{HS_L}(z) = \frac{z - e^{(-1/fs\tau\kappa)}}{z - e^{(-1/fs\tau)}}$$

$$H_{HS_R}(z) = \frac{z - e^{-2/fs\tau}}{z - e^{(-2\gamma/fs\tau)}} \qquad (3)$$

$$\tau = a/2c$$

where fs = sampling rate, a = head radius, c = speed of sound, $\kappa = 1\text{-}A\sin(B\theta)$, $\gamma = 1+C\sin(D\theta)$, $0 \le \theta \le \pi/2$ and A, B, C, D are scaling coefficients.

With A=B=1, the model provides minimum gain (-15dB when properly scaled) to the shadowed (far) ear at an azimuth of θ=+90°, which increases to no gain (0dB) as the azimuth moves to θ=0°. It was subsequently determined that this was not the ideal model, and that the far-ear was maximally shadowed closer to θ=+60°. The far-ear head-shadow was adjusted (A and B coefficients) so that the gain dips at θ=+60° and then increases somewhat as the azimuth approaches θ=+90°.

In addition to compensating the far-ear, the near-ear head-shadow filter increases the gain to the near ear. The increase in gain for the near ear is maximum (+10dB when properly scaled) at θ=+90° decreasing to none (0dB) at θ=0°.

The inverse head-shadow model is derived from the above equations. It is simply the inverse of the head-shadow model, and is of the form:

$$H_{IHS_L}(z) = \frac{1}{H_{HS_L}(z)} = \frac{z - e^{(-1/fs\tau)}}{z - e^{(-1/fs\tau\kappa)}}$$

$$H_{IHS_R}(z) = \frac{1}{H_{HS_R}(z)} = \frac{z - e^{-2\gamma/fs\tau}}{z - e^{(-2/fs\tau)}} \qquad (4)$$

$$\tau = a/2c$$

where fs = sampling rate, a = head radius, c = speed of sound, $\kappa = 1\text{-}A\sin(B\theta)$, $\gamma = 1+C\sin(D\theta)$, $0 \le \theta \le \pi/2$ and A, B, C, D are scaling coefficients.

Figure A presents the frequency response of the inverse head-shadow model.

The model used of the ITD is given below as referenced in [4]:

$$\Delta\tau_{ITD} = \frac{a}{c}(\theta + \sin\theta)fs \qquad (5)$$

where fs = sampling rate, a = head radius and c = speed of sound.

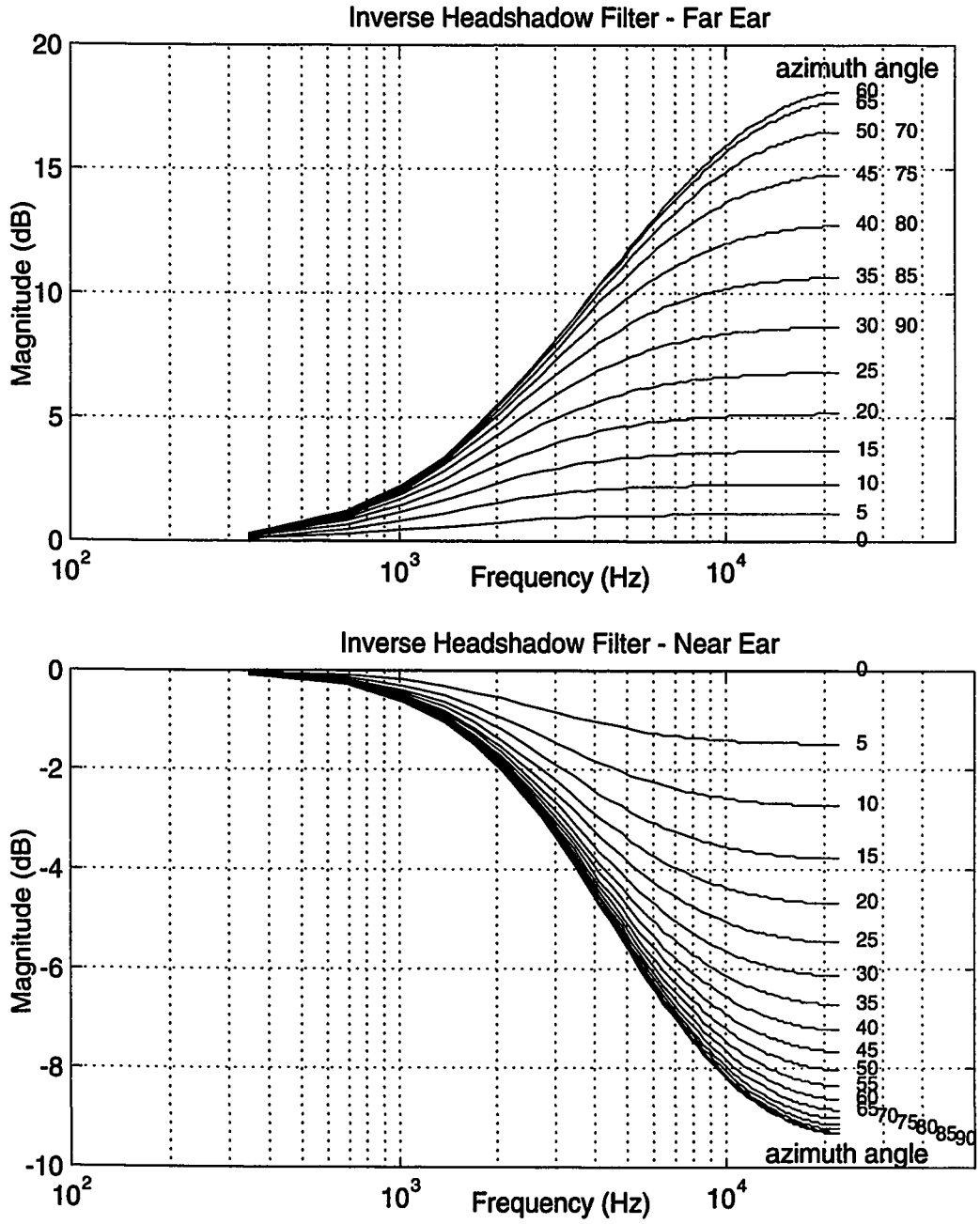


Figure A - Frequency Response of the Inverse Head-Shadow Model

Appendix B - Coordinate System

The coordinate system used for data collected in this thesis is a spherical, head-centered system, with the polar axis coinciding with the interaural axis (Figure A). The azimuth angle (θ) is restricted to the range $-90° < \theta < +90°$, while the elevation (ϕ) ranges through a full 360° circle. For a given azimuth angle, this system provides a basically constant ITD for all elevations.

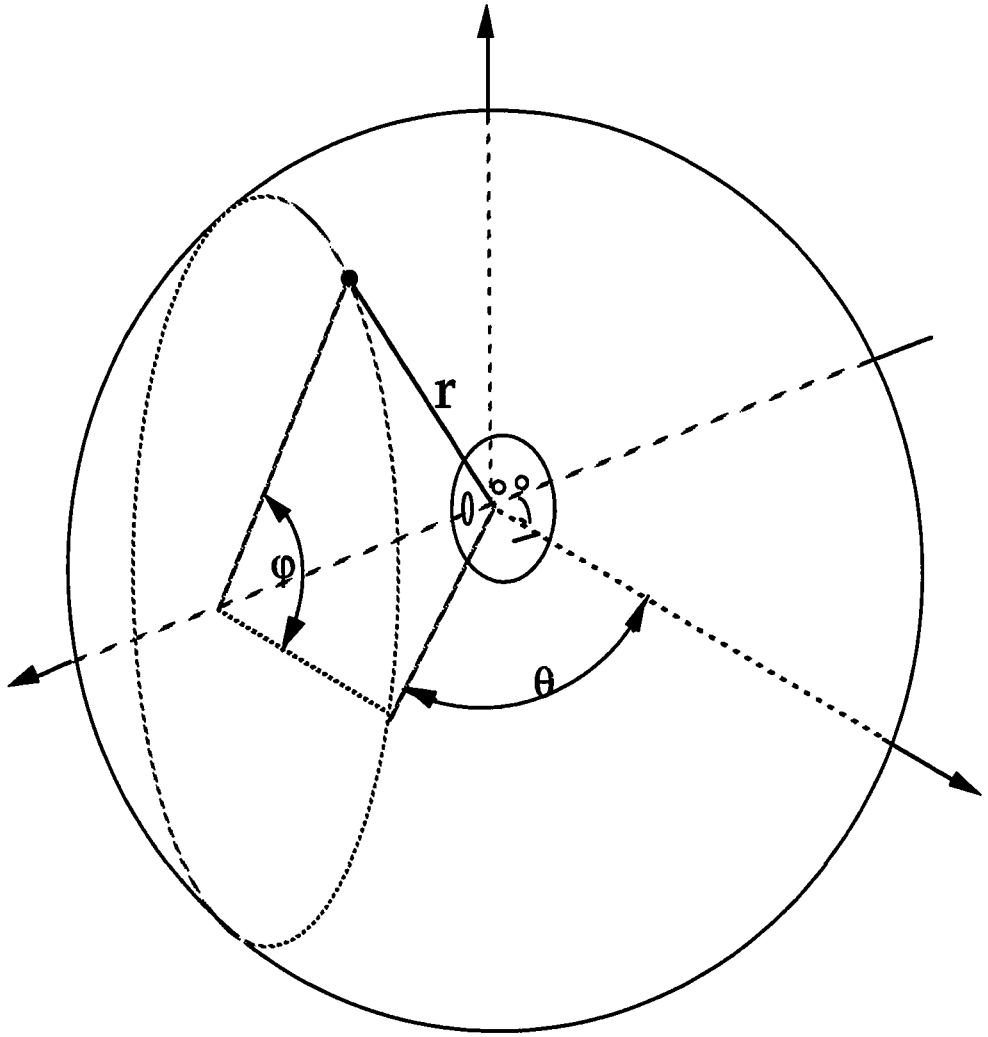


Figure B - Spherical Coordinate System
azimuth (θ), elevation (ϕ) and range(r)

Appendix C - Measurement Setup

A test fixture was designed to accommodate the polar coordinate system of Appendix A and is illustrated in Figure B. The fixture allows vertical movement from $-85° \leq \phi \leq +90°$ in the frontal plane by adjusting a pivoting pole, with the pivot point along the subject's interaural axis. Because the pivoting pole slightly interferes with the base pole at -90°, it was decided to limit the data on subsequent tests to ±80°.

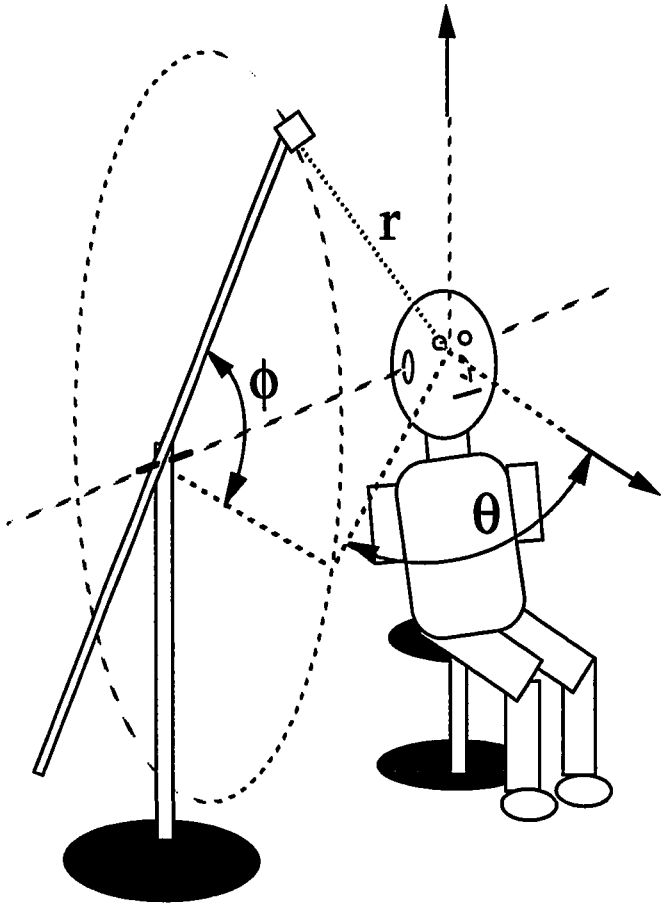


Figure C - Spherical Measurement Setup
azimuth (θ), elevation (ϕ) and range(r)

Appendix D - Matlab Code

A portion of the Matlab code used to generate the model presented in this thesis is given below. For the sake of brevity, only the core synthesis code is presented, while the support code (such as interface, playback, etc.) and alternate synthesis versions are omitted.

Headfilter3.m

This function contains the head-shadow model described in the text. A stereo array is passed to the function along with the azimuth angle, sampling rate, and a switch that determines whether the head-shadow or inverse head-shadow is applied.

```
function outcolumns = headfilter3(az,incolumns,invertflag,fs)
%function outcolumns = headfilter3(az,incolumns,invertflag,fs)
% positive az in radians toward right
% apply headshadow filter, or its inverse if nonzero third arg
% now set up to use 0.9*sin(az) where Duda proposed sin(az)
% positive az is shift away from column 1, toward column 2 if any

if size(incolumns,2) > 2, fprintf(2, 'too many columns in headfilter'), end

if nargin < 4, fs = 44100; end
if nargin < 3, invertflag=0; end

a = 0.0875;     % head radius in meters
c = 343;        % speed of sound in meters per second
tau = a/(2*c);  % pole time constant

feedback = [1, -exp(-1/(fs*tau))];  % feedback coefficients

%fudge the azimuth, first moving into range -pi/2 to pi/2, then upping it:
if az>pi/2, az = pi-az; end
if az<-pi/2, az = -pi+az; end
az1 = az;
az = 1.5*az;  % puts a dimple near the interaural poles, max near 70 degrees.

azx = -0.9*sin(az);
forward = [1, -exp(-1/(fs*tau*(1+azx)))];
gain = sum(forward)/sum(feedback);
```

```
if invertflag,
        outcolumns = filter(feedback,forward,incolumns(:,1))*gain;
else,
        outcolumns = filter(forward,feedback,incolumns(:,1))/gain;
end

if size(incolumns,2) == 1, return, end

az = az1;
azx = 2.2*sin(az); % steeper correction factor here
% second column has negative the azimuth of first:
tau = tau/2; % zero at 2.6 kHz, pole higher
forward = [1, -exp(-1/(fs*tau))];          % zero an octave above left pole
feedback = [1, -exp(-1/(fs*tau/(1+azx)))]; % movable pole
gain = sum(forward)/sum(feedback);

if invertflag,
        outcolumns = [outcolumns, filter(feedback,forward,incolumns(:,2))*gain];
else,
        outcolumns = [outcolumns, filter(forward,feedback,incolumns(:,2))/gain];
end
```

Pinna_model_5.m

This is the script to synthesize the HRIR model and convolve it with noise. For the listening tests, this code was reduced to a number of smaller functions, and the convolution was done "on the fly" through a GUI. One important thing to note is that the script generates the delays in a continuous fashion which are later put into a discrete format by splitting the delay between adjacent samples.

```
% pinna_model_5.m
% script to model pinna and shoulder echos
%
% the azimuth is fixed at 55°
%
% this version has three pinna echoes, as opposed to 2
% and does not use the original smoothing filter
% and instead uses the B'worth 1st order

clear all
hold off
echo off
```

```
fs=44100;
az=55;

load pb_align.mat %loads aligned hrir
hrir=hrir(:,37:72);

el=(-85:5:90)';

direct=ones(length(el),1);

% these numbers adjust the range of the sinusoidal response taken
% 0 to pi for tweak=1, etc.
%rd's
%D1=.85;
%D2=.35;

%pb's
D1=1;
D2=.5;

%nh's
%D1=1;
%D2=.5;

C1=0.5;
C2=0.5;

%shoulder coeff's
As=44;
Bs=12;
Cs=0.5;
Ds=0.6;

%generate delays
pinna_1=( cos(C1*az*pi/180)*sin(D1*(el+90)*pi/180) );
pinna_2=( cos(C2*az*pi/180)*sin(D2*(el+90)*pi/180) );
shld=round( As*(sin(Cs*az*pi/180)*cos(Ds*(el+90)*pi/180)) + Bs);


%put into "samples" quantity
A1=1;           %1st echo slope
B1=2;           %1st echo offset
A2=5;           %2nd echo
B2=4;
A3=5;           %3rd echo
B3=7;
A4=5;           %4th echo
B4=11;
A5=5;           %5th echo
B5=13;
```

```
% round down delay to nearest sample
pinna_1_fix=fix(pinna_1*A1 + B1);
pinna_2_fix=fix(pinna_2*A2) + B2;
pinna_3_fix=fix(pinna_2*A3) + B3;
pinna_4_fix=fix(pinna_2*A4) + B4;
pinna_5_fix=fix(pinna_2*A5) + B5;

%get the amount truncated due to rounding
pinna_1_mag=pinna_1*A1-(pinna_1_fix-B1);
pinna_2_mag=pinna_2*A2-(pinna_2_fix-B2);
pinna_3_mag=pinna_2*A3-(pinna_3_fix-B3);
pinna_4_mag=pinna_2*A4-(pinna_4_fix-B4);
pinna_5_mag=pinna_2*A5-(pinna_5_fix-B5);

% scale magnitude coeff's for echoes (a+b+c+d+e+f = 1)
a=1;
b=.5;
c=-1;
d=.5;
e=-.25;
f=.25;

% generate array with all the echoes
% and split and scale delays between adjacent samples
synthrir=zeros(size(hrir,1)+1,size(hrir,2));
for el=-85:5:90,
        idx=(el+85)/5+1;
        synthrir(shld(idx),idx)=0.2*cos(((idx-1)/35)*90*pi/180);
        synthrir(direct(idx),idx)=a;
        synthrir(pinna_1_fix(idx),idx)=b*(1-pinna_1_mag(idx));
        synthrir(pinna_1_fix(idx)+1,idx)=b*pinna_1_mag(idx);
        synthrir(pinna_2_fix(idx),idx)=c*(1-pinna_2_mag(idx));
        synthrir(pinna_2_fix(idx)+1,idx)=c*pinna_2_mag(idx);
        synthrir(pinna_3_fix(idx),idx)=d*(1-pinna_3_mag(idx));
        synthrir(pinna_3_fix(idx)+1,idx)=d*pinna_3_mag(idx);
        synthrir(pinna_4_fix(idx),idx)=e*(1-pinna_4_mag(idx));
        synthrir(pinna_4_fix(idx)+1,idx)=e*pinna_4_mag(idx);
        synthrir(pinna_5_fix(idx),idx)=f*(1-pinna_5_mag(idx));
        synthrir(pinna_5_fix(idx)+1,idx)=f*pinna_5_mag(idx);
end

%smoothing filter
[b,a]=butter(1,3000/(fs/2));
synthrir=filters(b,a,synthrir);

% truncate to 32 samples long
synthrir=synthrir(1:32,:);
hrir=hrir(1:32,:);

% up-sample for smoother image
for i=1:36,
```

```
        synthrir4(:,i)=interp(synthrir(:,i),4);
        hrir4(:,i)=interp(hrir(:,i),4);
end

for i=1:36,
        synthrir(:,i)=decimate(synthrir4(:,i),4);
        hrir(:,i)=decimate(hrir4(:,i),4);
end

tm=128/(4*fs/1000);
hrir4=hrir4/max(max(abs(hrir4)));
synthrir4=synthrir4/max(max(abs(synthrir4)));

% plot images of the measured and modeled data

figure(1)
imagesc([],[0 tm],hrir4);
%grid on
title('HRIR amplitude vs. time vs. elevation angle, near ear, azimuth=55°');
xlabel('Elevation Angle (degrees)')
ylabel('Time (ms)')
colormap(gray)
%colormap('jet')
colorbar

figure(2)
imagesc([],[0 tm],synthrir4)
%grid on
colormap(gray)
%colormap('jet')
colorbar
title('Model HRIR amplitude vs. time vs. elevation angle, near ear, azimuth=55°');
xlabel('Elevation Angle (degrees)')
ylabel('Time (ms)')
drawnow

break

% the remainder of the script convolves the model with noise and adds
% headshadow and ITD info

synthrir=[synthrir synthrir];
fprintf('[Insert left & right ear headshadow into synthrir...');
for i=1:36,
        outcolumns = headfilter3(pi*az/180,[synthrir(:,i) synthrir(:,i+36)],0,fs);
        synthrir(:,i)=outcolumns(:,1);
        synthrir(:,i+36)=outcolumns(:,2);
        fprintf('.');
        if i==25
                fprintf('\n');
        end
```

```
end
fprintf(']\n');

nsamples=size(synthrir,1);

fprintf('[Generating Gaussian white noise ...]\n');
duration = .5;              % duration in seconds
nduration = fix(duration*fs); % number of samples for duration
g_noise=randn(1,nduration);
g_noise=g_noise/(6*max(abs(g_noise))); %scale the noise, use 6
[b,a]=butter(1,1500/(fs/2),'high'); %HPF simulates ff response
g_noise=filter(b,a,g_noise);   % conv w/ meas'd ff response

%adjust nduration for the convolved signal
nduration=length(g_noise)+size(synthrir,1)-1;

%convolve the noise with the model
fprintf('[Convolving Gaussian noise with HRTF ...');
model_hrir_snd=zeros(nduration,72);
for i=1:72,
        model_hrir_snd(:,i)=conv(g_noise,synthrir(:,i)')';
        fprintf('.');
        if i==25
                fprintf('\n');
        end
end
fprintf(']\n');


fprintf('[Adding artificial ITD ...');
a = 0.09;     % head radius in meters
c = 343;      % speed of sound in meters per second
ndelay=round( (a/c)*(az*pi/180 + sin(az*pi/180))*fs); %samples for ITD
for i=1:36,
        model_hrir_snd(ndelay+1:nduration,i)=model_hrir_snd(1:nduration-ndelay,i);
        model_hrir_snd(1:ndelay,i)=zeros(ndelay,1);
        fprintf('.');
end;
fprintf(']\n');
```